Detecting changes in the extremal behavior of time series

Dissertation

zur

Erlangung des Grades Dr. rer. pol.
der Wirtschaftswissenschaftlichen Fakultät
der Universität Duisburg-Essen

vorgelegt von
Yannick Peer Hoga

Mai 2016

Datum der mündlichen Prüfung: 18.10.2016

Erstgutachter: Prof. Dr. Christoph Hanck
Zweitgutachter: Prof. Dr. Rüdiger Kiesel

Frequently used symbols and abbreviations:

$\Omega \neq \emptyset$	:	non-empty set
$\lvert\Omega\rvert$	:	cardinality of Ω
$\mathbb{N}$	$:=$	$\{1, 2, \ldots\}$
$\mathbb{Z}$	$:=$	$\{\ldots, -1, 0, 1, \ldots\}$
$\mathbb{R}$	:	set of real numbers
$\mathbb{R}^+$	$:=$	$(0, \infty)$
$(\mathbb{R}, \mathcal{B})$	:	$\mathbb{R}$ equipped with the Borel σ-field
I_A	:	$\Omega \to \{0,1\}\,,\; I_A(\omega) = \begin{cases} 1 & , \omega \in A; \\ 0 & , \omega \notin A. \end{cases}$
$a \vee b$	$:=$	$\max(a, b)$
$a \wedge b$	$:=$	$\min(a, b)$
$X_{1:n} \leq \ldots \leq X_{n:n}$	:	order statistics of r.v.s $X_1, \ldots X_n$
$\stackrel{\mathcal{D}}{=}$	:	equality in distribution
$\lfloor \cdot \rfloor$	:	$\mathbb{R} \to \mathbb{N}$, $x \mapsto \max\{n \in \mathbb{Z} : n \leq x\}$
$\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$	$:=$	$\left\{ f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}) \,\middle\vert\, \lVert f\rVert_p := \left(\int \lvert f\rvert^p \,\mathrm{d}\mu \right)^{1/p} < \infty \right\}$
$a(n) \sim b(n)$	$:\Leftrightarrow$	$\lim\limits_{n\to\infty} \frac{a(n)}{b(n)} = 1;$
$f(n) = \mathcal{O}(g(n))$	$:\Leftrightarrow$	$\limsup\limits_{n\to\infty} \frac{\lvert f(n)\rvert}{\lvert g(n)\rvert} < \infty;\quad f, g : \mathbb{N} \to \mathbb{R}$
$f(n) = o(g(n))$	$:\Leftrightarrow$	$\lim\limits_{n\to\infty} \frac{\lvert f(n)\rvert}{\lvert g(n)\rvert} = 0;\quad f, g : \mathbb{N} \to \mathbb{R}$
$f(n) \approx g(n)$	$:\Leftrightarrow$	$f(n) = \mathcal{O}(g(n))$ und $g(n) = \mathcal{O}(f(n))\,;$
$N(\mu, \sigma^2)$	:	normal distribution with expectation μ and variance σ^2
r.v.	:	random variable
i.i.d.	:	independent, identically distributed
a.s.	:	almost surely
d.f.	:	distribution function
cts.	:	continuous

Contents

Preface

With the rising recognition of outliers, the study of heavy-tailed phenomena has received more and more research attention over the past few decades. One of the first to recognize the presence of fat tails in financial markets was Mandelbrot (1963). Ever since, heavy-tailed distributions have been detected in a wide ranges of fields - from climatology (e.g., wind speeds) and insurance markets (e.g., insurance claim sizes) to teletraffic data (e.g., file size distributions); see Resnick (2007); Beirlant *et al.* (2004). The arguably most popular measure for the heavy-tailedness of a distribution is the tail index, usually denoted by α. It determines the maximal moment exponent (which is α), tail asymptotics of the distribution and the asymptotic behavior of sums and maxima.

While the theory for heavy tails was initially developed for i.i.d. data, some progress towards extending those methods to dependent sequences has been made (see Drees, 2008, for an overview). This was required as (e.g.) financial data typically exhibit conditionital heteroscedasticity (Engle, 1982) and hence cannot credibly be viewed as independently and identically distributed. Most authors from the extreme value theory literature considering dependent data made use of mixing concepts (e.g., Hsing, 1991; Drees, 2000), although more general dependence concepts have also been investigated (Hill, 2010).

Usually mixing conditions were assumed along with strict stationarity, which implies constancy of the tail index. However, since the work of Quintos *et al.* (2001) there is mounting evidence that the tail index is liable to change over time. The tests of Quintos *et al.* (2001) and their subsequent generalizations exclusively made use of the Hill (1975) estimator. It is the purpose of Chapter 2 to show how other tail index estimators may be used. It will turn out that using the moments-ratio estimator of Daníelsson *et al.* (1996), which was shown in Wagner and Marsh (2004) to compare favorably with the Hill (1975) estimator for ARCH-data, usually leads to more powerful tests.

However, while the tail index itself is of interest mainly in insurance, the closely-related quantity value-at-risk (VaR), which is a quantile of the distribution (usually of the returns of some speculative asset), is of more importance in the financial industry as it determines capital requirements (Daníelsson, 2011). Hence, it is natural to try to extend the tests from Chapter 2 to a setting where constancy of VaR is tested. This is done in Chapter 3. Quite generally, the tests developed in Chapter 3 are more powerful than those from Chapter 2, because the extreme quantile estimators used in the former tests also take differences in scale into account.

The tests proposed in Chapters 2 and 3 are of a retrospective nature - given a data set a decision can be made of whether a structural change occurred or not. Yet in practice, data usually arrive ‘online’ and each time more observations become available a decision has to made if there was a break. This leads to so called sequential (or also: online) monitoring procedures proposed in Chapter 4, which was co-authored by Dominik Wied (TU Dortmund and Universität zu Köln).

Finally, Chapter 5 investigates how to test for a change in the extremal dependence of the components of a random vector $\boldsymbol{V}_i = (X_i, Y_i)$. Structural break tests for measures of dependence over the whole real line have been investigated for quite general time series models (e.g., Aue *et al.*, 2009; Wied *et al.*, 2012), while those where dependence is measured only in the extremal region of both components are restricted to an i.i.d. setting (Bücher *et al.*, 2015). An extension of the results in Bücher *et al.* (2015) to β-mixing data is provided in Chapter 5.

Preceding the main results in Chapters 2 to 5, is a chapter giving a necessarily brief introduction to some of the main concepts used thereafter. Finally, Chapter 6 concludes.

* * *

My thanks goes first to my advisor Christoph Hanck. He taught me a lot about writing a good paper - from starting out with an interesting empirical question to stating the contribution of the final paper clearly. He also introduced me to the world of academia and guided me patiently through the process of writing cover letters, letters to referees, referee reports, etc. For all of this and the freedom he granted me, I owe him a great debt.

My thanks also goes to the anonymous referees of my papers that make up Chapters 2 through 5, whose comments lead to substantial improvements of those papers in content as well as presentation. The present dissertation has benefited a lot from peer review.

Last, but certainly not least, I owe the deepest gratitude to my wife, the silent co-author. Without her, no line of this dissertation would have been written, as I would have attended to the every need of my two little sons. She is doing a tremendous job.

1 Technical preliminaries

The workhorse of modern change point analysis (and many more areas in statistics and probability) is weak convergence in the space of so called càdlàg-functions. The classic reference on this is Billingsley (1968). Section 1.1 is devoted to an introduction of the space D of càdlàg-functions, i.e., right-continuous functions with left-hand limits. The acronym stems from the French designation 'continue à droite, limites à gauche'. Having introduced a 'suitable' metric in D, we continue by discussing weak convergence first on general metric spaces (Section 1.2.1) and then specialize these results to D (Section 1.2.2). In these chapters we will use concepts from topology, which are discussed briefly and succinctly in Billingsley (1968, Appendix).

Initially, most weak convergence results in D were developed for i.i.d. data, e.g., Donsker's invariance principle. (As an aside, Donsker's proof was not quite correct due to problems of measurability of functionals of discontinuous stochastic processes. A fix was provided by Skorohod, who introduced (one such) 'suitable' metric on D, such that convergence in D to a continuous function with respect to that metric is equivalent to convergence in the usual sup norm (see Dudley, 1999, Sec. 1.1 & Notes to Sec. 1.1).) In view of applications, we need to allow for more general dependence structures. We do so by appealing to the concept of β-mixing introduced in Section 1.3.

As much of the estimators upon which we base our change point tests are motivated by extreme value theory, we also provide some main results of this literature in Chapter 1.4. There, we also highlight the connection between extreme value theory and the theory of regularly varying functions.

1.1 The space D

This section draws heavily on Billingsley (1968, Chapter 3.14). The above mentioned space D is defined more precisely as follows:

Definition 1.1. The space $D = D\,[0,1]$ of **càdlàg functions** is defined as

$$D := \Bigg\{ x : [0,1] \to \mathbb{R} \;\Bigg|\; \forall\, t \in [0,1) : \; x\,(t+) := \lim_{s \downarrow t} x\,(s) \text{ exists, } x\,(t+) = x\,(t)\,,$$
$$\text{and } \forall\, t \in (0,1] : \; x\,(t-) := \lim_{s \uparrow t} x\,(s) \text{ exists, } x\,(1) = x\,(1-) \Bigg\}.$$

Remark 1.1. (i) $D[a,b]$ for $-\infty < a < b < \infty$ is defined analogously.

(ii) For $x \in D$: $\sup_{t\in[0,1]} |x(t)| < \infty$.

For a function $x \in C := C[0,1]$, the space of continuous and real functions on $[0,1]$, the modulus of continuity is defined as

$$w_x(\delta) := \sup_{|s-t|<\delta} |x(s) - x(t)|.$$

It will be useful to have an equivalent in D (see Theorem 1.9 below). For $x \in D$ and $T_0 \subset [0,1]$ define

$$w_x(T_0) := \sup_{s,t\in T_0} |x(s) - x(t)|. \tag{1.1}$$

For $\delta \in (0,1)$ set

$$w'_x(\delta) := \inf\left\{ \max_{i=1,\dots,r} w_x([t_{i-1}, t_i)) \;\middle|\; 0 = t_0 < \ldots < t_r = 1,\ \min_{i=1,\dots,r} (t_i - t_{i-1}) > \delta \right\}.$$

Remark 1.2. a) For $0 < \delta \leq 1$ and using (1.1) we can define the modulus of continuity of $x \in C$ as

$$w_x(\delta) = \sup_{0\leq t\leq 1-\delta} w_x([t, t+\delta))$$

b) Since for all $\delta < 1/2$ the interval $[0,1)$ can be split up into sub-intervals $[t_{i-1}, t_i)$ with $\delta < t_i - t_{i-1} \leq 2\delta$, we have

$$w'_x(\delta) \leq w_x(2\delta), \quad \text{if } \delta < \frac{1}{2}. \tag{1.2}$$

As already mentioned, we shall introduce a metric on D. To motivate this metric we first consider the space C. C equipped with $\rho : C \times C \to [0,\infty)$,

$$\rho(x,y) := \sup_{t\in[0,1]} |x(t) - y(t)| \tag{1.3}$$

is a metric space (C,ρ). Hence, two functions $x, y \in C$ are 'close' with respect to the metric ρ, if the graph of $x(t)$ can be obtained by that of $y(t)$ by a slight variation of the ordinate values, while keeping the abscissa fixed. The idea of the introduction of a metric on D is to also allow for uniformly small variations in abscissa values.

Those are generated by functions in

$$\Lambda := \left\{\lambda : [0,1] \to [0,1] \mid \lambda \text{ cts., strictly increasing, } \lambda(0) = 0, \lambda(1) = 1\right\}.$$

Definition 1.2. For $x, y \in D$ define the **Skorohod metric** $d : D \times D \to [0,\infty)$,

$$d(x,y) := \inf\left\{\varepsilon > 0 \;\middle|\; \exists \lambda \in \Lambda : \sup_{t\in[0,1]} \left|x(\lambda(t)) - y(t)\right| \le \varepsilon, \sup_{t\in[0,1]} \left|\lambda(t) - t\right| \le \varepsilon\right\}.$$

This metric induces a topology called the **Skorohod topology**. Denote by $\mathcal{D}$ the Borel-σ-field.

d, as just defined, is a metric on D (cf. Billingsley, 1968, p. 111). The uniformly small perturbations of the abscissa are captured by the requirement that

$$\sup_{t\in[0,1]} \left|\lambda(t) - t\right| \le \varepsilon, \tag{1.4}$$

i.e., λ is not too far from the identity function with respect to the metric ρ on C.

Remark 1.3. a) As usual, convergence in a metric space is defined as follows. Let $\{x_n\}_{n\in\mathbb{N}} \subset D$, $x \in D$. Then

$$\begin{aligned} x_n \underset{(n\to\infty)}{\overset{Skor}{\longrightarrow}} x \quad &:\Longleftrightarrow \quad \forall\, \varepsilon > 0\, \exists\, n_0 \in \mathbb{N} :\; d(x_n, x) \le \varepsilon \quad \forall\, n \ge n_0 \\ &\Longleftrightarrow \quad \exists \{\lambda_n\}_{n\in\mathbb{N}} \subset \Lambda : \sup_{t\in[0,1]} \left|x_n(\lambda_n(t)) - x(t)\right| \to 0, \\ &\qquad\qquad\qquad\qquad \sup_{t\in[0,1]} \left|\lambda_n(t) - t\right| \to 0. \end{aligned}$$

b) A mapping $f : (D,d) \to (\mathbb{R}, |\cdot|)$ is continuous in $x \in D$ if and only if the following implication holds

$$x_n \underset{(n\to\infty)}{\overset{Skor}{\longrightarrow}} x \quad \Longrightarrow \quad f(x_n) \underset{(n\to\infty)}{\longrightarrow} f(x).$$

Example 1.1. (i) $f : (D,d) \to (\mathbb{R}, |\cdot|)$, $f(x) := \sup_{t\in[0,1]} \left|x(t) - tx(1)\right|^2$ is a cts. mapping.

(ii) $w_{\cdot}(\delta) : (D,d) \to (\mathbb{R}, |\cdot|)$, $w_x(\delta) := \sup_{0<|t-s|<\delta} \left|x(t) - x(s)\right|$ is cts. in $x \in C \subset D$ for any $\delta \in (0,1)$.

The metric space (D,d) is however not complete (cf. Billingsley, 1968, p. 112). A metric that is defined slightly differently remedies this drawback by allowing for slightly different variations in abscissa values than in (1.4). We require the slope $\left(\lambda(t)-\lambda(s)\right)/(t-s)$ to be close to 1, which is the slope of the identity function. Or, put differently, we require the logarithm of $\left(\lambda(t)-\lambda(s)\right)/(t-s)$ to be close to 0. Hence, put

$$\|\lambda\| := \sup_{s\neq t}\left|\log\left(\frac{\lambda(t)-\lambda(s)}{t-s}\right)\right|.$$

Definition 1.3. For $x,y\in D$ define the **Billingsley-metric** $d_0 : D\times D\to[0,\infty)$,

$$d_0(x,y) := \inf\left\{\varepsilon>0 \,\middle|\, \exists\lambda\in\Lambda : \sup_{t\in[0,1]}\left|x(\lambda(t))-y(t)\right|\leq\varepsilon,\ \|\lambda\|\leq\varepsilon\right\}.$$

This metric induces a topology called the **Billingsley-topology**. Denote by $\mathcal{D}_0$ the Borel-σ-field.

d_0 is a metric on D (cf. Billingsley, 1968, p. 113) and we have

Theorem 1.1 (Billingsley, 1968, Theorem 14.1)**.** *The metrics d and d_0 are equivalent.*

Remark 1.4. a) It is well-known that equivalent metrics induce the same topology, i.e., a family of sets designated as open. Thus, the Borel-σ-fields $\mathcal{D}$ and $\mathcal{D}_0$ coincide and we simply write $\mathcal{D}$ in the following.

b) Further, due to the above theorem, a sequence in D converges with respect to the norm d if and only if it converges with respect to the norm d_0.

Remark 1.5. In this dissertation we will also have to deal with m-dimensional vectors $x=(x_1,\ldots,x_m)'$ of càdlàg functions in D. The space of these functions will be denoted by D^m. In D^m

$$d_0^m(x,y) := \max_{j\in\{1,\ldots,m\}}\left\{d_0(x_j,y_j)\right\}$$

defines a metric, that induces the product (Billingsley) topology. $D^m[a,b]$ for $-\infty<a<b<\infty$ is again defined analogously. For more details we refer to Davidson (1994, Chapter 29.5).

The metric space (D,d_0) is complete, as the following theorem shows.

Theorem 1.2 (Billingsley, 1968, Theorem 14.2)**.** *The metric space (D,d_0) is separable and complete.*

For convergence of probability measures defined on $(D,\mathcal{D})$ the so called projections play a key role (see Theorem 1.8 below) - for $t_1,\ldots,t_k\in[0,1]$ define the projection

$\pi_{t_1,\ldots,t_k}$ on D as follows:

$$\begin{aligned}\pi_{t_1,\ldots,t_k} : D &\to \mathbb{R}^k \\ x(\cdot) &\mapsto \left(x(t_1), \ldots, x(t_k)\right)\end{aligned}$$

Remark 1.6. $\pi_{t_1,\ldots,t_k}$ are $(\mathcal{D}, \mathcal{B}^k)$-measurable.

1.2 Weak convergence

1.2.1 Weak convergence in metric spaces

This section draws on Billingsley (1968, Chapter 1). Let S be a metric space equipped with metric ρ and Borel-σ-field $\mathcal{S}$, generated by the open (with respect to ρ) subsets in S. Hence, $(S, \mathcal{S})$ is a measure space. In Chapters 2 through 5 the space $(S, \mathcal{S}) = (D, \mathcal{D})$ will mainly be of interest.

Definition 1.4. Let $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$ and P be probability measures on $(S, \mathcal{S})$. We say that $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$ **converges weakly** to P (written $\mathrm{P}_n \underset{(n\to\infty)}{\Longrightarrow} \mathrm{P}$), if

$$\lim_{n\to\infty} \int f d\mathrm{P}_n = \int f d\mathrm{P} \quad \forall\, f \in C(S),$$

where $C(S) := \{f : S \to \mathbb{R} \mid f \text{ cts., bounded}\}$.

Remark 1.7. a) $f : S \to \mathbb{R}$ cts. in $x \in S \quad :\Longleftrightarrow$

$$\forall\, \varepsilon > 0\, \exists\, \delta > 0 : \ \rho(x_n, x) \leq \delta \Rightarrow \left| f(x_n) - f(x) \right| \leq \varepsilon.$$

b) The sequence $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$ from the above definition converges to a unique limit (cf. Billingsley, 1968, Theorem 1.3).

c) Note for the concept defined in the above definition that it only depends on the topology of S. In particular, two metrics inducing the same topology (e.g., two equivalent metrics, like d and d_0) generate the same classes $\mathcal{S}$ and $C(S)$ and hence lead to the same notion of weak convergence.

The following definition introduces an important concept.

Definition 1.5. (a) A probability measure P is **tight**, if

$$\forall\, \varepsilon > 0\, \exists\, K \subset S \text{ compact}: \quad \mathrm{P}(K) > 1 - \varepsilon.$$

(b) A family Π of probability measures is **tight**, if

$$\forall\, \varepsilon > 0\, \exists\, K \subset S \text{ compact}: \quad \mathrm{P}(K) > 1 - \varepsilon \quad \forall\, P \in \Pi.$$

Remark 1.8. a) A probability measure on a product space (e.g., D^m from Remark 1.5) is tight if and only if all marginal distributions on the component spaces are tight (cf. Billingsley, 1968, p. 41).

b) Consider the family of probability measures $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$ with P_n discrete putting mass one-half at 0 and one-half at n. Then it is easy to check that $\{\mathrm{P}_n\}$ is not tight, as it has 'mass escaping to infinity'.

An extensive characterization of weak convergence is given by the *Portmanteau theorem*:

Theorem 1.3 (Billingsley, 1968, Theorem 2.1)**.** *Let* $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$, P *be probability measures on* $(S,\mathcal{S})$. *Then the following are equivalent:*

(i) $\mathrm{P}_n \underset{(n\to\infty)}{\Longrightarrow} \mathrm{P}$*;*

(ii) $\lim_{n\to\infty} \int f d\mathrm{P}_n = \int f d\mathrm{P} \quad \forall\, f : S \to \mathbb{R}$ *uniformly cts., bounded;*

(iii) $\limsup_{n\to\infty} \mathrm{P}_n(F) \le \mathrm{P}(F) \quad \forall\, F \in \mathcal{S}$ *closed;*

(iv) $\limsup_{n\to\infty} \mathrm{P}_n(G) \ge \mathrm{P}(G) \quad \forall\, G \in \mathcal{S}$ *open;*

(v) $\lim_{n\to\infty} \mathrm{P}_n(A) = \mathrm{P}(A) \quad \forall\, A \in \mathcal{S}$ *mit* $\mathrm{P}(\partial A) = 0$.

Remark 1.9. Note for part *(v)* that $\partial A \in \mathcal{S}$, as ∂A is closed.

The theory of weak convergence can be reformulated as a theory of convergence in distribution, because of the one-to-one correspondence between probability measures and distributions of so called random elements (which we shall be interested in), a definition of which is given in

Definition 1.6. Let $(\Omega,\mathcal{A},\mathrm{P})$ be a probability space. A mapping $X : \Omega \to S$ is a **random element** it it is $(\mathcal{A},\mathcal{S})$-measurable.
X is also called a **random variable** if $(S,\mathcal{S}) = (\mathbb{R},\mathcal{B})$; a **random vector** if $(S,\mathcal{S}) = (\mathbb{R}^k,\mathcal{B}^k)$ $(k \ge 2)$; a **càdlàg process** if $(S,\mathcal{S}) = (D,\mathcal{D})$.

Corollary 1.1. *Let* $(\Omega,\mathcal{A},\mathrm{P})$ *be a probability space,* $X : \Omega \to D$ *a mapping. Then:*

$$X \text{ is a càdlàg process, i.e., } X^{-1}(\mathcal{D}) \subset \mathcal{A} \quad \Longleftrightarrow \quad \forall\, t \in [0,1] : \; X(t) \text{ is a real r.v., i.e., } X(t)^{-1}(\mathcal{B}) \subset \mathcal{A}.$$

Remark 1.10.

a) The above corollary shows that every càdlàg process is a stochastic process, i.e., a family $\{X_t\}_{t\in[0,1]}$ of real r.v.s.

b) The mappings of the form $X : \Omega \to D$ that will appear in the sequel are easily identified as càdlàg processes using this corollary.

Definition 1.7. The **distribution** of a random element X is denoted by P_X, i.e., the probability measure on $(S,\mathcal{S})$ defined by

$$\mathrm{P}_X(A) := \mathrm{P}\left(X^{-1}(A)\right), \quad A \in \mathcal{S}.$$

Definition 1.8. Let $\{X_n\}_{n\in\mathbb{N}}$ and X be random elements taking values in $(S,\mathcal{S})$.

1. $\{X_n\}_{n\in\mathbb{N}}$ **converges in distribution** to X (written $X_n \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} X$) if
$$\mathrm{P}_{X_n} \underset{(n\to\infty)}{\Longrightarrow} \mathrm{P}_X.$$
.
2. $\{X_n\}_{n\in\mathbb{N}}$ is **tight** if the family of probability measures $\{\mathrm{P}_{X_n}\}_{n\in\mathbb{N}}$ is tight.

Remark 1.11. The random elements in the above definition may all be defined on different probability spaces, as long as they take values in $(S,\mathcal{S})$.

In terms of random elements $\{X_n\}_{n\in\mathbb{N}}$ and X the Portmanteau-Theorem claims the equivalence of the following statements:

(i) $X_n \overset{\mathcal{D}}{\longrightarrow} X \ (n\to\infty)$;

(ii) $\lim_{n\to\infty} \mathbb{E}_n f(X_n) = \mathbb{E} f(X) \quad \forall\, f:\mathbb{R}\to\mathbb{R}$ unformly cts., bounded;

(iii) $\limsup_{n\to\infty} \mathrm{P}_n(X_n\in F) \le \mathrm{P}(X\in F) \quad \forall\, F\in\mathcal{S}$ closed;

(iv) $\limsup_{n\to\infty} \mathrm{P}_n(X_n\in G) \le \mathrm{P}(X\in G) \quad \forall\, G\in\mathcal{S}$ open;

(v) $\lim_{n\to\infty} \mathrm{P}_n(X_n\in A) = \mathrm{P}(X\in A) \quad \forall\, A\in\mathcal{S}$ with $\mathrm{P}(X\in\partial A)=0$.

Here, $\mathbb{E}_n f(X_n) := \int f(X_n)\,d\mathrm{P}_n$, $\mathbb{E} f(X) := \int f(X)\,d\mathrm{P}$.

We can also define the well-known concept of convergence in probability for random elements. Concretely:

Definition 1.9. Let $\{X_n\}_{n\in\mathbb{N}}$ be random elements on a probability space $(\Omega,\mathcal{A},\mathrm{P})$ taking values in the metric space (S,ρ). $\{X_n\}$ is said to **converge in probability** to $X\in S$ (written $X_n \underset{(n\to\infty)}{\overset{\mathrm{P}}{\longrightarrow}} X$) if
$$\mathrm{P}\left(\rho(X_n,X)\ge\varepsilon\right) \underset{(n\to\infty)}{\longrightarrow} 0 \quad \forall\,\varepsilon>0.$$

Theorem 1.4 (Billingsley, 1968, Theorem 4.1)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ be random elements on a probability space $(\Omega,\mathcal{A},\mathrm{P})$ taking values in a separable, metric space (S,ρ). Then:*

If $X_n \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} X$ and $\rho(X_n,Y_n) \underset{(n\to\infty)}{\overset{\mathrm{P}}{\longrightarrow}} 0$, then $Y_n \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} X$.

An important question is under which assumptions convergence in distribution is preserved. An answer is provided by the following

Theorem 1.5 (Billingsley, 1968, Theorem 5.1)**.** *Let* (S, ρ), (S', ρ') *be metric spaces equipped with the Borel* σ*-fields* $\mathcal{S}$, $\mathcal{S}'$ *and* $h : (S, \mathcal{S}) \to (S', \mathcal{S}')$ *a measurable mapping. Put* $D_h := \{x \in S \mid h(\cdot) \text{ discontinuous in } x\}$*. Then:*

If $\mathrm{P}_n \underset{(n\to\infty)}{\Longrightarrow} \mathrm{P}$ *and* $\mathrm{P}(D_h) = 0$*, then* $\mathrm{P}_n h^{-1} \underset{(n\to\infty)}{\Longrightarrow} \mathrm{P}h^{-1}$.

The following is also known as the *continuous mapping theorem.*

Corollary 1.2 (Billingsley, 1968, Corollary 1)**.** *Under the assumptions of Theorem 1.5 let* $\{X_n\}_{n\in\mathbb{N}}$, X *be random elements taking values in* $(S, \mathcal{S})$.
Íf $X_n \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} X$ *and* $\mathrm{P}(X \in D_h) = 0$*, then* $h(X_n) \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} h(X)$.

Remark 1.12. $D_h \in \mathcal{S}$, even if h is not measurable (cf. Billingsley, 1968, S. 225).

The following concept will help in formulating conditions for weak convergence of probability measures:

Definition 1.10. Let Π be a family of probability measures on $(S, \mathcal{S})$. Π is called **relatively compact** $:\Longleftrightarrow$

$$\forall \{\mathrm{P}_n\} \subset \Pi\, \exists \{\mathrm{P}_{n_k}\} \subset \{\mathrm{P}_n\}, \text{ a probability measure } Q \text{ on } (S, \mathcal{S}) : \mathrm{P}_{n_k} \underset{(k\to\infty)}{\Longrightarrow} Q.$$

The connection between relative compactness and tightness of a sequence of probability measures is clarified by the following two theorems, which together are known as *Prohorov's theorem.* In both theorems Π denotes a family of probability measures on a metric space $(S, \mathcal{S})$.

Theorem 1.6 (Billingsley, 1968, Theorem 6.1)**.** *If* Π *is tight, then* Π *is also relatively compact.*

Theorem 1.7 (Billingsley, 1968, Theorem 6.2)**.** *Let* S *be separable and compact. If* Π *is relatively compact, then* Π *is also tight.*

1.2.2 Weak convergence in D

This subsection draws on Billingsley (1968, Chapter 3.15).

Let P be a probability measure on $(D, \mathcal{D})$ and

$$T_{\mathrm{P}} := \{t \in [0,1] \mid \exists\, N_t \in \mathcal{D}, \mathrm{P}(N_t) = 0 : \pi_t \text{ cts. } \forall\, x \notin N_t\} \tag{1.5}$$

the set of points $t \in [0, 1]$ for which the projection π_t is cts. outside a P-null measure.

D equipped with d_0 is a separable and complete space (Theorem 1.2) and hence, according to Prohorov's theorem, a family of probability measures is relatively compact if and only if it is tight. This is a central element in the proof of the following important

Theorem 1.8 (Billingsley, 1968, Theorem 15.1)**.** *Let* $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$, P *be probability measures on* $(D, \mathcal{D})$*.*
If the sequence $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$ *is tight and*

$$\mathrm{P}_n\pi^{-1}_{t_1,\ldots,t_k} \underset{(n\to\infty)}{\Longrightarrow} \mathrm{P}\pi^{-1}_{t_1,\ldots,t_k} \quad \forall\, t_1,\ldots,t_k \in T_{\mathrm{P}},$$

then

$$\mathrm{P}_n \underset{(n\to\infty)}{\Longrightarrow} \mathrm{P}.$$

Remark 1.13. a) In Theorem 1.8 the converse is also true, i.e., if $\mathrm{P}_n \underset{(n\to\infty)}{\Longrightarrow} \mathrm{P}$, the familiy $\{\mathrm{P}_n\}$ is tight (since weakly converging probability measures are trivially relatively compact) and the *finite-dimensional distributions* $\mathrm{P}_n\pi^{-1}_{t_1,\ldots,t_k}$ for $t_1,\ldots,t_k \in T_{\mathrm{P}}$ converge (see Theorem 1.5).

b) The above theorem gives a widely-used 'recipe' to prove weak convergence: first, check convergence of the finite-dimensional distributions and, second, prove tightness of $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$.

In view of the second part of the above remark it is useful to have a more easily verified condition for (or even characterization of) tightness, which is given in the following

Theorem 1.9 (Billingsley, 1968, Theorem 15.2)**.** *A sequence of probability measures* $\{\mathrm{P}_n\}_{n\in\mathbb{N}}$ *on* $(D, \mathcal{D})$ *is tight* $\iff$

(i) $\forall\, \eta > 0\, \exists\, a > 0$*:*

$$\mathrm{P}_n \left\{ x \in D \,\middle|\, \sup_{t\in[0,1]} \left|x(t)\right| > a \right\} \le \eta \quad \forall\, n \ge 1; \tag{1.6}$$

(ii) $\forall\, \varepsilon, \eta > 0\, \exists\, \delta \in (0,1)\,, n_0 \in \mathbb{N}$*:*

$$\mathrm{P}_n \left\{ x \in D \,\middle|\, w'_x(\delta) \ge \varepsilon \right\} \le \eta \quad \forall\, n \ge n_0. \tag{1.7}$$

Theorems 1.8 und 1.9 can also be phrased in terms of sequences of càdlàg processes. Recall (1.5) and put for a càdlàg process X

$$T_X := T_{\mathrm{P}_X}. \tag{1.8}$$

Theorem 1.10. *A sequence* $\{X_n\}_{n\in\mathbb{N}}$ *of càdlàg processes is tight and satisfies*

$$(X_n(t_1), \ldots, X_n(t_k))' \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} (X(t_1), \ldots, X(t_k))' \quad \forall\, t_1, \ldots, t_k \in T_X$$

if and only if

$$X_n \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} X.$$

Theorem 1.11. *A sequence* $\{X_n\}_{n\in\mathbb{N}}$ *of càdlàg processes is tight* $\iff$

(i) $\forall\, \eta > 0\, \exists\, a > 0$*:*

$$\mathrm{P}\left\{\sup_{t\in[0,1]} |X_n(t)| > a\right\} \leq \eta \quad \forall\, n \geq 1; \tag{1.9}$$

(ii) $\forall\, \varepsilon, \eta > 0\, \exists\, \delta \in (0,1), n_0 \in \mathbb{N}$*:*

$$\mathrm{P}\left\{w'_{X_n}(\delta) \geq \varepsilon\right\} \leq \eta \quad \forall\, n \geq n_0. \tag{1.10}$$

Remark 1.14. Conditions (1.9) and (1.10) are often written compactly as

$$\begin{aligned} (1.9) \quad &\iff \quad \sup_{t\in[0,1]} |X_n(t)| = \mathcal{O}_P(1) \quad (n\to\infty); \\ (1.10) \quad &\iff \quad \forall\, \varepsilon > 0: \lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathrm{P}\left\{w'_{X_n}(\delta) \geq \varepsilon\right\} = 0. \end{aligned}$$

1.3 Sequences of mixing random variables

This section collects definitions and results from Bradley (1986, 2007). The latter reference provides an encyclopedic treatment of mixing conditions. We focus here on β-mixing, the concept used in Chapters 2 through 5. Indeed, there is a wealth of other mixing concepts, e.g., that of α-mixing and ρ-mixing.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$ two sub-σ-fields. Define

$$\beta(\mathcal{F}, \mathcal{G}) \quad := \quad \mathbb{E}\left[\sup_{F \in \mathcal{F}} \left|\mathrm{P}(G \mid \mathcal{F}) - \mathrm{P}(G)\right|\right],$$

where $\beta(\mathcal{F}, \mathcal{G})$ is well-defined if $\mathcal{G}$ is separable, i.e., there exist at most countably many events $A_1, A_2, \ldots \in \mathcal{A}$ with $\mathcal{G} = \sigma(\{A_1, A_2, \ldots\})$, and a regular conditional probability $\mathrm{P}(G \mid \mathcal{F})$ $(G \in \mathcal{G})$ (cf. Bradley, 2007, 3.22 Proposition). Note that (Bradley, 2007, p. 10)

$$\mathcal{G} \subset \mathcal{A} \text{ separable} \quad \Longleftrightarrow \quad \mathcal{G} = \sigma(X) \text{ for some r.v. } X.$$

The following more general definition does not require separability and the existence of a regular conditional probability (cf. Bradley, 2007, 3.3 Definitions). For $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$:

$$\beta(\mathcal{F}, \mathcal{G}) := \frac{1}{2} \sup \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} \left|\mathrm{P}(A_i \cap B_j) - \mathrm{P}(A_i)\,\mathrm{P}(B_j)\right| \right\},$$

where the supremum is taken over all partitions with $\sum_{i=1}^{I} A_i = \sum_{j=1}^{J} B_j = \Omega$, $A_i \in \mathcal{F}$, $B_j \in \mathcal{G}$.

In the remainder of this section we assume that $\{X_i\}_{i\in\mathbb{Z}}$ is a sequence of random variables defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and

$$\mathcal{F}_a^b := \sigma(X_i : \ a \le i \le b) \quad (-\infty \le a < b \le \infty)$$

is the σ-field generated by the respective r.v.s. As we are ultimately interested in general dependence concepts for r.v.s, we can now define:

Definition 1.11. $\{X_i\}_{i\in\mathbb{Z}}$ is called **β-mixing** if

$$\beta(n) := \sup_{k\in\mathbb{Z}} \beta(\mathcal{F}_{-\infty}^{k}, \mathcal{F}_{k+n}^{\infty}) \underset{(n\to\infty)}{\longrightarrow} 0$$

holds for the **β-mixing coefficients** $\beta(n)$.

Remark 1.15 (Bradley, 1986, pp. 170 and 173). (i) The mixing coefficients of a sequence of independent r.v.s satisfy

$$\beta(n) = 0 \quad \forall\, n \in \mathbb{N},$$

i.e., such a sequence is trivially β-mixing.

(ii) If $\{X_i\}_{i\in\mathbb{N}}$ is a 'singly-infinite' sequence one modifies Definition 1.11 to $\beta(n) := \sup_{k\in\mathbb{N}} \beta(\mathcal{F}_1^k, \mathcal{F}_{k+n}^\infty)$.

(iii) If $\{X_i\}_{i\in\mathbb{Z}}$ is strictly stationary, the β-mixing coefficients simplifiy to

$$\beta(n) = \beta(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty) = \mathbb{E}\left[\sup_{A\in\mathcal{F}_n^\infty} \left|\mathrm{P}(A \mid \mathcal{F}_{-\infty}^0) - \mathrm{P}(A)\right|\right] \quad \forall\, n \in \mathbb{N}.$$

(iv) The sequence $\{\beta(n)\}_{n\in\mathbb{N}}$ is obviously non-increasing in n.

(v) If $\{f_i : \mathbb{R} \to \mathbb{R}\}_{i\in\mathbb{Z}}$ is a sequence of measurable functions and $\{X_i\}_{i\in\mathbb{Z}}$ β-mixing with β-mixing coefficients $\beta(n)$, then the sequence $\{f_i(X_i)\}_{i\in\mathbb{Z}}$ is also β-mixing with β-mixing coefficients

$$\beta_f(n) \le \beta(n) \quad \forall\, n \in \mathbb{N}.$$

(vi) If $\{X_i\}_{i\in\mathbb{Z}}$ and $\{Y_i\}_{i\in\mathbb{Z}}$ are β-mixing sequences that are independent of each other, then the random vector $(X_i, Y_i)'$ is also β-mixing. Hence, the sequences of sums $X_i + Y_i$ and products $X_i \cdot Y_i$ are also β-mixing.

(vii) Items (i)-(iv) carry over word for word to (e.g.) ρ-mixing r.v.s.

Another possible definition using r.v.s is given by

Theorem 1.12 (Bradley, 2007, 3.30 Corollary)**.** *Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be metric spaces and X, Y random elements taking values in S_1, S_2 respectively. Then*

$$\beta(\mathcal{A}(X), \mathcal{A}(Y)) = \sup_{A\in\mathcal{S}_1\otimes\mathcal{S}_2} \left|P_{(X,Y)}(A) - P_X \otimes P_Y(A)\right|.$$

This theorem will be particularly useful in connection with the following lemma. For probability measures μ, ν on a measure space $(S, \mathcal{S})$ define the norm

$$\|\mu - \nu\| := \sup_{A\in\mathcal{S}} |\mu(A) - \nu(A)|.$$

Lemma 1.1 (Eberlein, 1984, Lemma 2). *Let $X_1, \ldots, X_n$ be random elements taking values in a measure space $(S, \mathcal{S})$. Let $\varepsilon > 0$ and suppose that*

$$\left\| \mathrm{P}_{(X_1,\ldots,X_n)'} - \mathrm{P}_{(X_1,\ldots,X_k)'} \otimes \mathrm{P}_{(X_{k+1},\ldots,X_n)'} \right\| \leq \varepsilon, \tag{1.11}$$

then

$$\left\| \mathrm{P}_{(X_1,\ldots,X_n)'} - \mathrm{P}_{X_1} \otimes \ldots \otimes \mathrm{P}_{X_n} \right\| \leq \varepsilon(n-1).$$

This lemma will be particularly useful in connection with a suitable β-mixing condition, as by Theorem 1.12

$$\left\| \mathrm{P}_{(X_1,\ldots,X_n)'} - \mathrm{P}_{(X_1,\ldots,X_k)'} \otimes \mathrm{P}_{(X_{k+1},\ldots,X_n)'} \right\| = \beta(\sigma(X_1,\ldots,X_k), \sigma(X_{k+1},\ldots,X_n)).$$

If $\beta(\sigma(X_1,\ldots,X_k), \sigma(X_{k+1},\ldots,X_n))$ is bounded by some sequence $\beta(l_n)$, such that $\beta(l_n)(n-1) = o(1)$, then Lemma 1.1 shows that, as far as convergence in distribution is concerned, one may as well assume $X_1,\ldots,X_n$ to be independent. See Eberlein (1984, Sec. 3), Kim and Lee (2009, Proof of Lemma 6) or the proof of Theorem 2.4 below for applications.

1.4 Extreme value theory

There are quite a few monographs on classical extreme value theory, e.g., de Haan and Ferreira (2006); Resnick (2007); Beirlant *et al.* (2004). The following introduction is based on the former reference.

1.4.1 Limit distributions and domains of attraction

Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. r.v.s with distribution function F. While classical central limit theory is concerned with the asymptotic behavior of partial sums $\frac{1}{n}(X_1 + \ldots + X_n)$, extreme value theory studies the asymptotic behavior of sample extremes like $\max(X_1, \ldots, X_n)$ or $\min(X_1, \ldots, X_n)$. Note that $\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n)$, so that it suffices to develop the theory for sample maxima.

Setting $x^* := \sup\{x \in \mathbb{R} : F(x) < 1\}$ ($\sup \mathbb{R} := \infty$), it is obvious that

$$\max(X_1, \ldots, X_n) \underset{(n\to\infty)}{\xrightarrow{\mathrm{P}}} x^*,$$

since

$$\begin{aligned}
\mathrm{P}\left(\max(X_1, \ldots, X_n) \le x\right) &= \mathrm{P}\left(X_1 \le x, \ldots, X_n \le x\right) \\
&= F^n(x) \underset{(n\to\infty)}{\longrightarrow} \begin{cases} 0 & , x < x^*; \\ 1 & , x \ge x^*. \end{cases}
\end{aligned}$$

The partial sum analogon is the law of large numbers, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{(n\to\infty)}{\xrightarrow{\mathrm{P}}} \mathbb{E}X_1.$$

In order to obtain (non-trivial) convergence in distribution results for $\max(X_1, \ldots, X_n)$ (in analogy to the central limit theorem) a normalization is necessary:

Definition 1.12. A non-degenerate d.f. $G : \mathbb{R} \to \mathbb{R}$ is called **extreme value distribution**, if for sequences $\{a_n\}_{n\in\mathbb{N}} \subset (0, \infty)$ and $\{b_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ it is the limiting d.f. of

$$\frac{\max(X_1, \ldots, X_n) - b_n}{a_n}, \tag{1.12}$$

i.e.,

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \quad \forall\, x \in C_G, \tag{1.13}$$

where C_G denotes the set of continuity points of G.
The set of d.f.s F with (1.13) is called the **domain of attraction** of G, written $F \in \mathcal{D}(G)$.

The question which non-degenerate d.f.s G are possible limiting d.f.s in (1.13) is answered by the following

Theorem 1.13 (de Haan and Ferreira, 2006, Theorem 1.1.3)**.** *The class of extreme value distributions is given by*

$$\left\{G_\gamma\left(a\cdot+b\right) \mid a>0;\, b,\gamma\in\mathbb{R}\right\},$$

where for $x\in\mathbb{R}$

$$G_\gamma\left(x\right)=\exp\left(-\left(1+\gamma x\right)^{-1/\gamma}\right),\quad 1+\gamma x>0. \tag{1.14}$$

For $\gamma=0$ *the right-hand side of* (1.14) *is interpreted as a limit, i.e.,* $\exp\left(-e^{-x}\right)$*. The parameter* $\gamma\in\mathbb{R}$ *is called* ***extreme value index*** *and* $\alpha=1/\gamma>0$ *is called* ***tail index*** *if* $\gamma>0$*.*

Remark 1.16. (1) Note that in the central limit theorem,

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}X_i-\mathbb{E}X_1}{\sqrt{\operatorname{Var}\left(X_1\right)}}\underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}}\mathcal{N}\left(0,1\right),$$

the limiting distribution does not depend on the underlying d.f. of the random variables.

(2) Obviously for $a>0$, $b\in\mathbb{R}$

$$F\in\mathcal{D}\left(G_\gamma\left(a\cdot+b\right)\right)\quad\Longleftrightarrow\quad F\in\mathcal{D}\left(G_\gamma\left(\cdot\right)\right).$$

In this case we write $F\in\mathcal{D}\left(G_\gamma\right)$.

(3) Theorem 1.13 shows that the class of extreme value distributions can be parametrized, apart from the scale parameter a and the location parameter b, by the single parameter γ. The d.f.s for $\gamma>0$, $\gamma=0$ und $\gamma<0$ have quite different characteristics:

a) $\gamma>0$: $G_\gamma\left(x\right)<1\ \forall x\in\mathbb{R}$; $1-G_\gamma\left(x\right)\sim\gamma^{-1/\gamma}x^{-1/\gamma}$ $(x\to\infty)$, i.e., the distribution has a heavy right tail; moments of order higher than $1/\gamma$ do not exist, while those of smaller order do.

b) $\gamma=0$: $G_\gamma\left(x\right)<1\ \forall x\in\mathbb{R}$; $1-G_0\left(x\right)\sim e^{-x}$ $(x\to\infty)$, i.e., the distribution has a light right tail; moments of arbitrary order exist.

c) $\gamma<0$: $G_\gamma\left(x\right)=1\ \forall x\geq-1/\gamma$, i.e., the distribution has a short tail with right end-point $-1/\gamma$.

A famous result, stated in Theorem 1.14 below, relates convergence of (1.12) to a nondegenerate limit, i.e.

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) \quad \forall\, x \in \mathbb{R}:\ 1+\gamma x > 0,$$

to the functional form of the d.f. F of the X_i's. Before stating it, we require the following

Definition 1.13 (de Haan and Ferreira, 2006, Definition B.1.1)**.** A function $f : (\mathbb{R}^+, \mathcal{B}_{(0,\infty)}) \to (\mathbb{R}, \mathcal{B})$, that is eventually positive, is **regularly varying with index** $\alpha \in \mathbb{R}$ (written $f \in RV_\alpha$) if

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^\alpha \quad \forall\, x > 0. \tag{1.15}$$

A function is **slowly varying** if it is regularly varying with index $\alpha = 0$.

Remark 1.17. From the above definition the following characterization is obvious:

$$f \in RV_\alpha \quad \Longleftrightarrow \quad f(x) = x^\alpha l(x), \qquad \text{where } l \in RV_0$$

Theorem 1.14 (de Haan and Ferreira, 2006, Theorem 1.2.1)**.** *The d.f.* $F : \mathbb{R} \to [0,1]$ *is in the domain of attraction of* G_γ *(*$\gamma \in \mathbb{R}$*) if and only if*

1. *For* $\gamma > 0$*:* $x^* := \sup\{x \in \mathbb{R} \mid F(x) < 1\} = \infty$ *and*

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma} \quad \forall\, x > 0.$$

2. *For* $\gamma < 0$*:* $x^* := \sup\{x \in \mathbb{R} \mid F(x) < 1\} < \infty$ *and*

$$\lim_{t\downarrow 0} \frac{1-F(x^*-tx)}{1-F(x^*-t)} = x^{-1/\gamma} \quad \forall\, x > 0.$$

3. *For* $\gamma = 0$*:* $x^* := \sup\{x \in \mathbb{R} \mid F(x) < 1\} \le \infty$ *and*

$$\lim_{t\uparrow x^*} \frac{1-F(t+xf(t))}{1-F(t)} = e^{-x} \quad \forall\, x \in \mathbb{R},$$

where f *is a suitable positive function.*

In typical financial applications estimates of $\alpha = 1/\gamma$ are in the range from 2 to 4 (Resnick, 2007). Hence, we will focus on positive γ in the following. If $\gamma > 0$, there are some useful equivalents of condition 1. in the above theorem to be found in de Haan and Ferreira (2006, Sec. 1.2). For instance, it is equivalent to the requirements that $F(x) < 1$ for all $x < \infty$, $\int_1^\infty (1 - F(x))/x\mathrm{d}x < \infty$ and

$$\lim_{t\to\infty} \frac{\int_t^\infty (1 - F(x))\frac{\mathrm{d}x}{x}}{1 - F(t)} = \gamma. \tag{1.16}$$

If we set

$$U(t) := F^{-1}\left(1 - \frac{1}{t}\right), \quad t > 1,$$

where F^{-1} is the left-continuous inverse of F, so that $U(t)$ is the $(1 - 1/t)$-quantile of F, then

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma$$

is also equivalent to 1.

Remark 1.18. For a r.v. $X > 0$ the d.f. F has support in $\mathbb{R}^+$. Then for $\gamma > 0$ according to the above theorem

$$F \in \mathcal{D}(G_\gamma) \quad \Longleftrightarrow \quad \bar{F} \in RV_{-1/\gamma},$$

which demonstrates the connection between extreme value theory and the theory of regularly varying functions.

1.4.2 The Hill estimator

As mentioned in the previous section financial data frequently exhibit heavy tails with some extreme value index $\gamma > 0$. Hence it is apt to introduce the arguably most popular estimator for a positive extreme value index - the Hill (1975) estimator.

Replacing t by $X_{n-k:n}$ and $F(\cdot)$ by the empirical d.f. $F_n(\cdot)$, relation (1.16) suggests the following estimator of γ:

$$\begin{aligned}
\widehat{\gamma}_H &:= \frac{\int_{X_{n-k:n}}^\infty (1 - F_n(x))\frac{\mathrm{d}x}{x}}{1 - F_n(X_{n-k:n})} \\
&= \frac{n}{k}\int_1^\infty \left(1 - F_n(xX_{n-k:n})\right)\frac{\mathrm{d}x}{x} \qquad (1.17)\\
&= \frac{1}{k}\sum_{i=1}^n \int_1^\infty I_{\left\{X_i > xX_{n-k:n}\right\}}\frac{\mathrm{d}x}{x} \\
&= \frac{1}{k}\sum_{i=0}^k \int_{X_{n-i:n}/X_{n-k:n}}^\infty \frac{\mathrm{d}x}{x}
\end{aligned}$$

$$= \frac{1}{k} \sum_{i=0}^{k} \log(X_{n-i:n}) - \log(X_{n-k:n})$$

which is the estimator proposed by Hill (1975). Consider (1.17) and write

$$\frac{n}{k} \left(1 - F_n(xX_{n-k:n})\right) = \frac{1}{k} \sum_{i=1}^{n} I_{\left\{X_i > xX_{n-k:n}\right\}} =: PSP.$$

This suggests central limit theory for the Hill estimator may be based on weak convergence results for the partial sum process PSP. This will be the approach taken in Chapter 2 under β-mixing conditions. In fact, a wide range of tail index estimators can be written as functionals of PSP, which explains the generality of the results in that chapter.

The Hill (1975) estimator however remains the probably most studied tail index estimator. It was investigated under α-mixing conditions in Hsing (1991) and under near-epoch dependence in Hill (2010). While its popularity stems mainly from certain optimality properties for i.i.d. data (e.g., Csörgő *et al.*, 1985), optimality for dependent data has, to the best of our knowledge, not been shown. Simulation evidence for ARCH-data in Wagner and Marsh (2004) suggests that the Hill (1975) estimator may no longer be the wisest choice in the presence of conditional heteroscedasticity.

2 Change point tests for the tail index of β-mixing random variables

The tail index as a measure of tail thickness provides information that is not captured by standard volatility measures. It may however change over time. Currently available procedures for detecting those changes for dependent data (e.g., Quintos *et al.*, 2001) are all based on comparing Hill (1975) estimates from different subsamples. We derive tests for a wide class of other tail index estimators. The limiting distribution of the test statistics is shown not to depend on the particular choice of the estimator, while the assumptions on the dependence structure allow for sufficient generality in applications. A simulation study investigates empirical sizes and powers of the tests in finite samples.

2.1 Motivation

The tail index of a distribution is of great importance in statistics, in particular in extreme value theory. It determines the limit distribution of the (suitably normalized) sample maximum and minimum. Also, the tail index determines the existence of higher-order moments and consequently is used as a measure for the thickness of the tail of a distribution. As such it is of interest in fields as diverse as finance, hydrology and internet-traffic engineering, where heavy tails are frequently encountered in real data. Moreover, it is important to know if the tail index of a time series has changed at some point during the observation period, since ignoring such a change can have negative consequences. For example, being unaware of a change to thicker tails of financial returns may lead to avoidable losses due to inadequate risk management, or, if tails vary from thicker to thinner, foregone profits because too much capital is put aside as a cushion against extreme losses. Indeed, there is empirical evidence that such changes do occur for many time series (Quintos, Fan and Phillips, 2001; Galbraith and Zernov, 2004; Werner and Upper, 2004).

The tail index is superior to other volatility measures, like the variance, when measuring volatility in at least the following two respects: It only captures the behavior of the distribution in the tail, upon which interest in, e.g., financial risk management frequently centers. This is by definition not the case for the variance, suggesting that there is information in the tail index that is not present in other volatility measures, which Werner and Upper (2004) also found indications for in empirical work.

Secondly, the variance as a measure of tail behavior is only available if second moments exist. Well-known empirical results show that this is not always the case (e.g., Resnick, 2007, Figs. 4.12 and 4.15). The tail index, in contrast, does not require the existence of p-th moments for some $p > 0$.

Much research has been devoted to tail index estimation, see, e.g., Drees (1998a,b) for some general results in the i.i.d. case and Drees (2000) for the dependent case. But estimating the tail index from a sample $X_1, \dots, X_n$ assumes (often implicitly) homogeneity in the tail index, which might not be warranted. A test of this assumption is useful for at least two reasons: First, in the case of an undetected break in the tail index from, say, α_1 to $\alpha_2 > \alpha_1$, where tails get lighter after the break, most tail index estimators will consistently estimate $1/\alpha_1$ (see Theorem 2.3 below), suggesting a heavier tail for the post-break period. In the above example of financial returns this would lead to excessive conservatism. Second, the tail behavior depends in a very sensitive way on the tail index. E.g., for a Student's t_2-distribution (where the tail index equals the degrees of freedom) the 99.9%-quantile is 22.3 and for a t_1-distribution the same quantity is 318. Combined with the first reason this suggests that an undetected break in the sample would lead to a wrong tail index estimate and hence a very misleading picture of the tail behavior.

Tests for a change in the tail index at a known breakpoint have been available for some time (Koedijk, Schafgans and de Vries, 1990). So called recursive, rolling and sequential tests for an unknown break point in the tail index were first proposed by Quintos *et al.* (2001) for i.i.d. and GARCH(1,1) data. These tests were subsequently investigated by Kim and Lee (2011) to cover strictly stationary, β-mixing random variables. All these tests are based on comparing Hill (1975) estimates of different subsamples. However, many other, possibly better (in a mean-squared error sense), estimators exist, e.g., Figure 1 in de Haan and Peng (1998) for the i.i.d. case and the simulation results in Wagner and Marsh (2004) for ARCH-type data. Very recently, under 'heteroscedastic extremes', Einmahl *et al.* (2016) allowed for other estimators to be used, although they only focused on the Hill (1975) estimator. Our first main contribution is to show that a vast range of tail index estimators is covered under their and (equivalently) our scheme, while, unlike Einmahl *et al.* (2016), allowing for dependent data, which is crucial for most real-world applications. Previously, consistency of change point tests for the tail index has only been proved under independence (Quintos *et al.*, 2001; Kim and Lee, 2009) or not at all (Einmahl *et al.*, 2016; Kim and Lee, 2011). The second main contribution is to demonstrate consistency under dependence. Further, we show that if there is a single tail index break in the sample, tail index estimators will still converge weakly, though with a different limit distribution. This result might be of independent interest.

A simulation study investigates whether gains in power can be achieved in a change point context by using other tail index estimators than Hill's covered by our framework. For instance for ARCH data 'there is a tendency of the Hill estimator to overestimate small tail indices and to underestimate large tail indices' (Wagner and Marsh, 2004, p. 3), which may lead to poor power properties of the change point tests based on the Hill estimator, as confirmed by our simulations. Another problem with the application of change point tests for the tail index to ARCH-models is that to the best of our knowledge, all currently available tests (cf. Quintos *et al.*, 2001) are derived using standard-normally distributed innovations, which in empirical work is often not credible (e.g., Aguilar and Hill, 2015, Fig. 2). This issue is also addressed in this chapter by allowing for, e.g., t-distributed innovations, which are also used in the simulations. Furthermore, our simulations reveal that the problem of nonmonotonic power, i.e., power not increasing monotonically in the distance from the null, is most severe for the Hill estimator.

The generality of our results rests on the insight that a wide range of tail index estimators can be written as functionals of the (slightly adapted) sequential tail empirical process. Allowing for dependence requires consistent estimates of the asymptotic variance of the tail index estimator. With the exception of Drees (2003), such estimators have so far only been considered for very specific dependence structures and tail index estimators, e.g., in Quintos *et al.* (2001) (for the Hill estimator and GARCH(1,1) data) and Chan *et al.* (2013) (for a moment-type estimator and AR(1) data with ARCH innovations). We propose a consistent variance estimator for the tail index estimators we consider under weak conditions on the dependence of the time series observations, which might be of independent interest.

The main results are stated in Section 2.2. Simulation evidence is presented in Section 2.3 and the proofs are relegated to Section 2.4.

2.2 Main results

This section is organized as follows: Subsection 2.2.1 introduces basic notation and the main assumptions that will be used throughout. It also gives examples of linear and nonlinear models for which these assumptions have been verified. Subsection 2.2.2 introduces some of the estimators that can be used under our scheme and states convergence results under the null. Results under a one-break alternative are stated in Subsection 2.2.3.

2.2.1 Preliminaries

Consider stationary r.v.s $\{X_i\}_{i\in\mathbb{N}}$ defined on some probability space $(\Omega, \mathcal{A}, P)$. Let F be the d.f. of X_1, where $1-F$ is assumed to be regularly varying with parameter $-\alpha < 0$ (written $1-F \in RV_{-\alpha}$), i.e.,

$$\frac{1-F(ty)}{1-F(t)} \underset{(t\to\infty)}{\longrightarrow} y^{-\alpha} \qquad \forall\, y > 0, \tag{2.1}$$

where α is called the tail index of X_1. If we define

$$U(t) := F^{\leftarrow}\left(1 - \frac{1}{t}\right), \qquad t > 1,$$

as the $(1 - 1/t)$-quantile, where $^{\leftarrow}$ denotes the left-continuous inverse, then (2.1) is equivalent to

$$\frac{U(ty)}{U(t)} \underset{(t\to\infty)}{\longrightarrow} y^{\gamma} \qquad \forall\, y > 0, \tag{2.2}$$

with $\gamma = 1/\alpha > 0$ the extreme value index (cf. de Haan and Ferreira, 2006). We will use both notations, γ and α.

Remark 2.1. If $\{X_i\}$ are i.i.d., then by the well-known Fisher-Tippett theorem the extreme value index γ determines (apart from a location and scale parameter) the possible limiting d.f.s of

$$\frac{\max(X_1, \ldots, X_n) - b_n}{a_n} \qquad (a_n > 0, b_n \in \mathbb{R}), \tag{2.3}$$

namely $G_{\gamma}(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right)$, $1 + \gamma x > 0$.

In the sequel, $k = k_n \in \mathbb{N}$ with $k \leq n - 1$ will be an intermediate sequence, i.e.,

$$k \underset{(n\to\infty)}{\longrightarrow} \infty \qquad \text{and} \qquad \frac{k}{n} \underset{(n\to\infty)}{\longrightarrow} 0,$$

controlling the number of “extremely large” observations used in the estimation of the tail index. For $t - s \geq 1/n$ and $y \in [0, 1]$ set

$$X_k(s, t, y) := \Big(\lfloor k(t-s)y \rfloor + 1\Big)\text{-th largest value of } X_{\lfloor ns \rfloor + 1}, \ldots, X_{\lfloor nt \rfloor}. \tag{2.4}$$

For clarity of exposition we will sometimes write $X_{1:n} \leq \ldots \leq X_{n:n}$ for the order statistics of $X_1, \ldots, X_n$. The dependence concept used here is that of β-mixing. Recall that a sequence of random variables $\{X_i\}_{i\in\mathbb{N}}$ is β-mixing iff

$$\beta(l) := \sup_{m\in\mathbb{N}} \mathbb{E}\left[\sup_{A\in\mathcal{F}_{m+l+1}^{\infty}} \left|\mathrm{P}(A|\mathcal{F}_1^m) - \mathrm{P}(A)\right|\right] \underset{(l\to\infty)}{\longrightarrow} 0,$$

where $\mathcal{F}_m^{\infty} := \sigma(X_m, X_{m+1}, \ldots)$ and $\mathcal{F}_l^m := \sigma(X_l, \ldots, X_m)$ are the σ-algebras generated by the respective r.v.s.

If it is in doubt whether all $X_1, \ldots, X_n$ have the same extreme value index $\gamma_1 = \ldots = \gamma_n$, it is important for reasons detailed in the motivation to test the following hypothesis:

$$\begin{aligned} \mathcal{H}_0 : &\quad \gamma_1 = \ldots = \gamma_n \qquad \text{versus} \\ \mathcal{H}_1 : &\quad \text{Not } \mathcal{H}_0. \end{aligned} \tag{2.5}$$

We now state our main assumptions that will be maintained throughout.

(A1) $\{X_i\}_{i\in\mathbb{N}}$ is a strictly stationary β-mixing process with continuous marginals and mixing coefficients $\beta(\cdot)$, such that

$$\lim_{n\to\infty} \frac{n}{r_n}\beta(l_n) + \frac{r_n}{\sqrt{k}}\log^2(k) = 0$$

for sequences $\{l_n\}_{n\in\mathbb{N}} \subset \mathbb{N}, \{r_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$ tending to infinity with $l_n = o(r_n)$, $r_n = o(n)$.

(A2) There exists a function $r(\cdot,\cdot)$, s.t. for all $x, y \in [0, y_0 + \delta]$ $(\delta > 0)$

$$\lim_{n\to\infty} \frac{n}{r_n k} \operatorname{Cov}\left(\sum_{i=1}^{r_n} I_{\left\{X_i > U\left(\frac{n}{kx}\right)\right\}}, \sum_{j=1}^{r_n} I_{\left\{X_j > U\left(\frac{n}{ky}\right)\right\}} \right) = r(x, y). \tag{2.6}$$

(A3) For some constant $C > 0$

$$\frac{n}{r_n k}\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U\left(\frac{n}{ky}\right) < X_i \le U\left(\frac{n}{kx}\right)\right\}} \right]^4 \le C(y - x) \qquad \forall\, 0 \le x < y \le y_0 + \delta,\ n \in \mathbb{N}.$$

(A4) There exist $\rho < 0$ and a function $A(\cdot)$ that is eventually positive or negative with $\lim_{t\to\infty} A(t) = 0$, s.t.

$$\lim_{t\to\infty} \frac{\frac{U(ty)}{U(t)} - y^\gamma}{A(t)} = y^\gamma \frac{y^\rho - 1}{\rho} \qquad \forall\, y > 0,$$

where $\sqrt{k}A(n/k) \to 0$ as $n \to \infty$.

Remark 2.2. (a) Conditions **(A1)**, **(A2)** and **(A3)** are discussed in some detail in Drees (2000, 2003) and Rootzén (2009). They are almost identical to conditions $(\widetilde{\text{C1}})$, $(\widetilde{\text{C2}})$ and $(\widetilde{\text{C3*}})$ in Drees (2000). **(A4)** is a standard second-order condition (used in, e.g., Einmahl *et al.*, 2016) that controls the speed of convergence in (2.2). It is slightly stronger than Drees' (2000) corresponding condition (3.5), which can be seen from de Haan and Ferreira (2006, Theorem 2.3.9).

(b) If **(C2)** holds for k_n and $k_{n,\lambda_j} \sim \lambda_j k_n$ ($\lambda_j \in (0,1)$), $j = 1, 2$, then (cf. Drees, 2003, p. 629)

$$r(tx, ty) = tr(x, y), \tag{2.7}$$

which will simplify the expressions for the asymptotic variances of the estimators we consider.

Conditions **(A1)-(A4)** relax some assumptions of previous tests. For instance, our scheme covers a wide range of short-memory processes (see Rootzén, 2009, Section 4, for an overview) in contrast to Einmahl *et al.* (2016), where only independent r.v.s are considered. Allowing for dependence is essential as, under dependence, the limit distribution of the test statistic considered in Einmahl *et al.* (2016, Corollary 2) will be scaled by some (dependence-structure dependent) factor. Note that the presence of 'heteroscedastic extremes' introduced in Einmahl *et al.* (2016) does not influence the limit behavior of their test statistic. Next, heavy-tailed innovations for ARCH(1)-processes are allowed under our conditions (and not in that of Quintos *et al.*, 2001); see also Remark 2.8.

The next two examples, taken from Drees (2000, Section 4) and Drees (2003, Subsections 3.1 & 3.2), give specific models where **(A1)-(A3)** have been verified. While the first-order condition in (2.2) is satisfied for both examples, the second-order condition **(A4)** has not yet been verified.

Example 2.1 (Linear model)**.** Consider stationary $\{X_i\}_{i\in\mathbb{N}}$ with representation

$$X_i = \sum_{j=0}^{\infty} \Psi_j Z_{i-j}, \quad i \in \mathbb{N},$$

where $\{Z_i\}_{i\in\mathbb{Z}}$ are i.i.d., $\Psi_0 = 1$ without loss of generality (w.l.o.g.) and $|\Psi_j| = \mathcal{O}\left(\tau^j\right)$, $j \to \infty$, for some $\tau \in (0,1)$. If for F_Z the d.f. of Z_1 we have $1 - F_Z \in RV_{-\alpha}$ ($\alpha > 0$) and some further conditions hold, then **(A1)-(A3)** hold for sequences $k = k_n$ satisfying

$$\log^2(n)\log^4(\log n) = o(k) \quad \text{and} \quad k = o\left(n/\log(n)\right). \tag{2.8}$$

In that case, the X_i also have tail index α.

Example 2.2 (Nonlinear model)**.** Consider a squared ARCH(1)-process $\left\{X_i^2\right\}_{i\in\mathbb{N}}$

$$X_i^2 = \left(\alpha_0 + \alpha_1 X_{i-1}^2\right) Z_i^2, \quad i \in \mathbb{N},$$

where $\alpha_0, \alpha_1 > 0$ and $\{Z_i\}_{i\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} (0,1)$. If Z_1 satisfies the following moment conditions for some $\kappa, \xi > 0$

$$\mathbb{E}\log(\alpha_1 Z_1^2) < 0, \quad \mathbb{E}(\alpha_1 Z_1^2)^{\kappa} = 1, \quad \mathbb{E}(\alpha_1 Z_1^2)^{\kappa+\xi} < \infty, \quad \mathbb{E}(\alpha_0 Z_1^2)^{\kappa+\xi} < \infty, \tag{2.9}$$

then conditions **(A1)-(A3)** were shown to hold for sequences $k = k_n$ satisfying

$$\log^2(n)\log^4(\log n) = o(k) \quad \text{and} \quad k = o\left(n^{2\rho/(2\rho+1)}\right) \quad \text{for some } \rho > 0.$$

The tail index $\alpha = \kappa > 0$ of (the strictly stationary) X_i^2 is determined by the moment condition $\mathbb{E}\left(\alpha_1 Z_i^2\right)^\alpha = 1$. Hence, α can be changed either by varying α_1 or the distribution of Z_i. Note that light-tailed Z_i, e.g., $Z_i \sim \mathcal{N}(0,1)$, lead to heavy tails in X_i, which is not true for Example 2.1.

2.2.2 Results under the null

The generality of our approach rests on a (weighted) weak convergence result for

$$F_n(s,t,y) := \frac{1}{\lfloor k(t-s)\rfloor} \sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor} I_{\left\{X_i > yX_k(s,t,1)\right\}}.$$

A wide range of tail index estimators can be written as functionals of $F_n(s,t,y)$. E.g., the Hill estimator $\widehat{\gamma}_H(0,1)$ based on the full sample $X_1,\ldots,X_n$ can be written as

$$\widehat{\gamma}_H(0,1) := \frac{1}{k}\sum_{i=0}^{k} \log\left(\frac{X_{n-i:n}}{X_{n-k:n}}\right) = \int_1^\infty F_n(0,1,y)\frac{\mathrm{d}y}{y}; \tag{2.10}$$

see also Examples 2.3-2.5 below. In a first step we will establish weighted convergence of the *sequential tail empirical process*

$$\sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i > yU(n/k)\right\}} - y^{-1/\gamma}t\right\}.$$

Then a variant of this result for $F_n(0,t,y)$ will be used to investigate weak convergence of (suitably normalized) generic tail index estimators based on subsamples $X_{\lfloor ns\rfloor+1},\ldots,X_{\lfloor nt\rfloor}$

$\widehat{\gamma}(s,t)$ that can be written as functionals of $F_n(s,t,y)$.

Under **(A1)-(A4)** it will be possible to derive the limiting distributions of the test statistics (where $\widehat{\sigma}^2_{\widehat{\gamma},\gamma} \in \{\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{nor}}, \widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}}\}$ is defined in Theorem 2.2 below)

$$
\begin{aligned}
Q_{\mathrm{rec}} &:= \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \sup_{t\in[t_0,1-t_0]} \left\{ t\sqrt{k}\left[\widehat{\gamma}(0,t)-\widehat{\gamma}(0,1)\right]\right\}^2;\\
Q^{\leftarrow}_{\mathrm{rec}} &:= \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \sup_{t\in[t_0,1-t_0]} \left\{ (1-t)\sqrt{k}\left[\widehat{\gamma}(t,1)-\widehat{\gamma}(0,1)\right]\right\}^2;\\
Q_{\mathrm{seq}} &:= \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \sup_{t\in[t_0,1-t_0]} \left\{ t(1-t)\sqrt{k}\left[\widehat{\gamma}(0,t)-\widehat{\gamma}(t,1)\right]\right\}^2;\\
Q_{\mathrm{rol}} &:= \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \sup_{t\in[t_0,1-t_0]} \left\{ t_0\sqrt{k}\left[\widehat{\gamma}(t,t+t_0)-\widehat{\gamma}(0,1)\right]\right\}^2;
\end{aligned}
\tag{2.11}
$$

for the testing problem (2.5), namely (see Corollary 2.2 below)

$$
\begin{aligned}
Q^{(\leftarrow)}_{\mathrm{rec}} &\underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \sup_{t\in[t_0,1-t_0]} \left\{W(t)-tW(1)\right\}^2,\\
Q_{\mathrm{seq}} &\underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \sup_{t\in[t_0,1-t_0]} \left\{W(t)-tW(1)\right\}^2,\\
Q_{\mathrm{rol}} &\underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \sup_{t\in[t_0,1-t_0]} \left\{\left[W(t+t_0)-W(t)\right]-t_0W(1)\right\}^2,
\end{aligned}
\tag{2.12}
$$

where $W(\cdot)$ denotes a standard Brownian motion. The general form of the test statistics in (2.11) is taken from Quintos *et al.* (2001). We have modified Q_{seq} slightly by including the factor $(1-t)$. Without it Q_{seq}, by construction, would be more likely to detect a change in the tail index at the end of the observation period, where t is large, than towards the beginning, which may not be desirable.

We assume throughout that $t_0 \in (0, 1/2)$. In our framework t_0 and $(1-t_0)$ denote the time before and after which the change is not allowed to occur. Since all tests allow to take t_0 arbitrarily close to zero, this does not impose a serious restriction. Further, by choosing t_0 closer to $1/2$ one can incorporate prior knowledge of the change point location in the tests, which, as unreported simulations for $Q^{(\leftarrow)}_{\mathrm{rec}}$ show, leads to higher power. It is easy to verify that asymmetric intervals à la $[t_0, t_1]$, $t_1 \in [t_0, 1)$, over which the supremum is taken in (2.12) and (2.11) are also possible, lending more flexibility to the incorporation of prior beliefs.

The weighted convergence result stated in the next theorem is fundamental to our approach (see also Remark 2.3 (b) below). For this, define non-negative, continuous weight functions $q(\cdot)$, similarly as in Drees (2000, Eq. (1.3)), as functions satisfying

$$
\inf_{y>\vartheta} q(y) > 0 \quad \forall\, \vartheta > 0 \qquad \text{and} \qquad y^{\nu}\left|\log y\right|^{\mu} = \mathcal{O}\left(q(y)\right), \quad y \downarrow 0, \tag{2.13}
$$

for some $\nu \in [0, 1/2)$, $\mu \in \mathbb{R}$ or $\nu = 1/2$, $\mu > 1/2$. Then we may prove

Theorem 2.1. *Suppose* **(A1)**-**(A4)** *hold and* $q(\cdot)$ *satisfies* (2.13). *Then for some* $\widetilde{\delta} > 0$*, under a Skorohod construction,*

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}-\widetilde{\delta}}} \frac{1}{q\left(y^{-1/\gamma}\right)} \left| \sqrt{k} \begin{pmatrix} \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>yU(n/k)\right\}} - y^{-1/\gamma}t \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{n} I_{\left\{X_i>yU(n/k)\right\}} - y^{-1/\gamma}(1-t) \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{\lfloor n(t+t_0)\rfloor} I_{\left\{X_i>yU(n/k)\right\}} - y^{-1/\gamma}t_0 \end{pmatrix} - \begin{pmatrix} W(t, y^{-1/\gamma}) \\ W(1, y^{-1/\gamma}) - W(t, y^{-1/\gamma}) \\ W(t+t_0, y^{-1/\gamma}) - W(t, y^{-1/\gamma}) \end{pmatrix} \right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0, \quad (2.14)$$

where $\{W(t,y)\}$ *is a continuous zero-mean Gaussian process with covariance function*

$$\operatorname{Cov}\left(W\left(t_1,y_1\right),W\left(t_2,y_2\right)\right) = \min\left(t_1,t_2\right) r\left(y_1,y_2\right).$$

A slightly modified version of (2.14), where $U(n/k)$ is replaced by an appropriate empirical counterpart, will be more convenient for our purposes. This results in a change of the limiting processes.

Corollary 2.1. *Suppose* **(A1)**-**(A4)** *hold for some* $y_0 \geq 1$ *and* $q(\cdot)$ *satisfies* (2.13). *Then, under a Skorohod construction,*

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}}} \frac{1}{q\left(y^{-1/\gamma}\right)} \left| \sqrt{k} \begin{pmatrix} t\left(F_n(0,t,y) - y^{-1/\gamma}\right) \\ (1-t)\left(F_n(t,1,y) - y^{-1/\gamma}\right) \\ t_0\left(F_n(t,t+t_0,y) - y^{-1/\gamma}\right) \end{pmatrix} - \begin{pmatrix} W(t,y^{-1/\gamma}) - y^{-1/\gamma}W(t,1) \\ W(1,y^{-1/\gamma}) - W(t,y^{-1/\gamma}) - y^{-1/\gamma}\left[W(1,1)-W(t,1)\right] \\ W(t+t_0,y^{-1/\gamma}) - W(t,y^{-1/\gamma}) - y^{-1/\gamma}\left[W(t+t_0,1)-W(t,1)\right] \end{pmatrix} \right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0, \quad (2.15)$$

where $\{W(t,y)\}$ *is as in Theorem 2.1.*

Since the above asymptotic results become stronger the smaller $q(\cdot)$ is, it would in fact suffice to only consider the case $\nu = 1/2$ in (2.13).

In the following three examples we will demonstrate how the convergence result of Corollary 2.1 can be used to establish joint convergence of $\sqrt{k}(\widehat{\gamma}(0,t)-\gamma, \widehat{\gamma}(t,1)-\gamma, \widehat{\gamma}(t,t+t_0)-\gamma)^T$.

Example 2.3 (WLS estimator)**.** Consider the class of weighted least squares (WLS) estimators of the tail index

$$\widehat{\gamma}_{WLS}(0,1) := \sum_{j=1}^{k} \int_{(j-1)/k}^{j/k} J(s)\mathrm{d}s \log(X_{n+1-j:n})$$

and with a finite-sample correction

$$\widehat{\gamma}_{\widetilde{WLS}}(0,1) := \frac{\sum_{j=1}^{k} \int_{(j-1)/k}^{j/k} J(s)\mathrm{d}s \log(X_{n+1-j:n})}{\sum_{j=1}^{k} \int_{(j-1)/k}^{j/k} J(s)\mathrm{d}s \log(k/j)}$$

discussed in Csörgő and Viharos (1998), where the weighting function $J(\cdot)$ satisfies

(W1) $\int_0^1 J(s)\mathrm{d}s = 0$,

(W2) $J(\cdot)$ is non-increasing and continuous on $[0,1]$,

(W3) $-\int_0^1 \log(s)J(s)\mathrm{d}s = 1$.

Proposition 2.1. *Suppose* **(C1)**-**(C4)** *hold for* $y_0 = 1$*. Then for* $J(\cdot)$ *satisfying* **(W1)**-**(W3)**

$$\sqrt{k}\begin{pmatrix} \widehat{\gamma}_{WLS}(0,t) - \gamma \\ \widehat{\gamma}_{WLS}(t,1) - \gamma \\ \widehat{\gamma}_{WLS}(t,t+t_0) - \gamma \end{pmatrix} \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma}_{WLS},\gamma} \begin{pmatrix} W(t)/t \\ (W(1)-W(t))/(1-t) \\ (W(t+t_0)-W(t))/t_0 \end{pmatrix} \tag{2.16}$$

in $D^3[t_0, 1-t_0]$*, where* $W(\cdot)$ *is a standard Brownian motion and*

$$\sigma^2_{\widehat{\gamma}_{WLS},\gamma} = \gamma^2 \int_0^1 \int_0^1 \frac{r(x,y)}{xy} J(x)J(y)\mathrm{d}x\mathrm{d}y. \tag{2.17}$$

Specifically, Csörgő and Viharos (1998) consider weight functions fulfilling **(W1)**-**(W3)**

$$J_\theta(s) := \frac{\theta+1}{\theta} - \frac{(\theta+1)^2}{\theta} s^\theta, \qquad s \in [0,1], \quad \theta > 0,$$

yielding estimators denoted by $\widehat{\gamma}_{CV_\theta}$, that possess certain optimality properties in a mean-squared error sense (cf. Csörgő and Viharos, 1998, Weight Theorem (ii)). Then, under (2.7), (2.17) simplifies to

$$\gamma^2 \int_0^1 \int_0^1 \frac{r(z,1)}{z} J_\theta(zy)J_\theta(y)\mathrm{d}y\mathrm{d}z = 2\frac{(\theta+1)^2}{2\theta+1}\gamma^2 \int_0^1 \frac{r(x,1)}{x} x^\theta \mathrm{d}x. \tag{2.18}$$

Remark 2.3. (a) Inclusion of the finite-sample correction does not change the convergence result in (2.16) (cf. Csörgő and Viharos, 1998, p. 18).

(b) The need for a weighted convergence result as in Corollary 2.1 can be seen most clearly from (2.61) in the proof of Proposition 2.1 below, where without weighting (i.e., $q \equiv 1$) the integral in that expression would not generally be finite.

Example 2.4 (Hill estimator)**.** As in the proof of Proposition 2.1 we will only derive weak convergence of $\widehat{\gamma}_H(0,t)$ from the first component of (2.15). Joint convergence as in (2.16) can again be obtained from (2.15). Check that similarly as in (2.10) we have $\widehat{\gamma}_H(0,t) = \int_1^\infty F_n(0,t,y)\frac{\mathrm{d}y}{y}$, such that by Corollary 2.1

$$\begin{aligned}
\sqrt{k}\left(\widehat{\gamma}_H(0,t) - \gamma\right) \quad &= \quad \sqrt{k}\int_1^\infty \left(F_n(0,t,y) - y^{-1/\gamma}\right)\frac{\mathrm{d}y}{y} \\
&\underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \quad \frac{1}{t}\int_1^\infty \left[W(t, y^{-1/\gamma}) - y^{-1/\gamma}W(t,1)\right]\frac{\mathrm{d}y}{y} \\
&= \quad \frac{\gamma}{t}\int_0^1 \left[W(t,u) - uW(t,1)\right]\frac{\mathrm{d}u}{u}.
\end{aligned}$$

Calculate covariances to obtain that

$$\frac{\gamma}{t}\int_0^1 \left[W(t,u) - uW(t,1)\right]\frac{\mathrm{d}u}{u} \overset{\mathcal{D}}{=} \sigma_{\widehat{\gamma}_H,\gamma}W(t)/t,$$

where $W(\cdot)$ denotes a standard Brownian motion and

$$\sigma^2_{\widehat{\gamma}_H,\gamma} = \gamma^2\int_0^1\int_0^1 \left\{\frac{r(x,y)}{xy} - \frac{r(x,1)}{x} - \frac{r(1,y)}{y} + r(x,y)\right\}\mathrm{d}x\mathrm{d}y \overset{(2.7)}{=} \gamma^2 r(1,1). \tag{2.19}$$

Example 2.5 (Moments ratio estimator)**.** We consider convergence of the moments ratio estimator based on the subsample $X_1, \ldots, X_{\lfloor nt\rfloor}$. Define, for $j = 1, 2$,

$$M_j(t) := \frac{1}{\lfloor kt\rfloor}\sum_{i=1}^{\lfloor kt\rfloor}\left(\log(X_{\lfloor nt\rfloor - i+1:\lfloor nt\rfloor}) - \log(X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor})\right)^j.$$

Then

$$\widehat{\gamma}_{MR}(0,t) := \frac{1}{2}\frac{M_2(t)}{M_1(t)} = \frac{1}{2}\frac{M_2(t)}{\widehat{\gamma}_H(0,t)}$$

is the so called moments ratio (MR) estimator of the tail index introduced by Daníelsson *et al.* (1996). One may verify that (cf. also the proof of Proposition 2.1)

$$M_1(t) = \int_1^\infty F_n(0,t,y)\frac{\mathrm{d}y}{y} \qquad \text{and} \qquad M_2(t) = \int_1^\infty F_n(0,t,y)2\log(y)\frac{\mathrm{d}y}{y}. \tag{2.20}$$

Then, under **(A1)-(A4)**, for $y_0 = 1$

$$\begin{aligned}
& \widehat{\gamma}_H(0,t)\cdot\sqrt{k}\left(\widehat{\gamma}_{MR}(0,t)-\gamma\right) \\
= \quad & \sqrt{k}\left(\frac{1}{2}\int_1^\infty F_n(0,t,y)2\log(y)\frac{\mathrm{d}y}{y} - \gamma\int_1^\infty F_n(0,t,y)\frac{\mathrm{d}y}{y}\right) \\
= \quad & \sqrt{k}\int_1^\infty F_n(0,t,y)\left[\log(y)-\gamma\right]\frac{\mathrm{d}y}{y} \\
= \quad & \int_1^\infty \sqrt{k}\left[F_n(0,t,y)-y^{-1/\gamma}\right]\left[\log(y)-\gamma\right]\frac{\mathrm{d}y}{y} \\
& \qquad\qquad +\sqrt{k}\int_1^\infty \left[\log(y)-\gamma\right]y^{-(1/\gamma+1)}\mathrm{d}y \\
\underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \quad & \frac{1}{t}\int_1^\infty \left[W(t,y^{-1/\gamma})-y^{-1/\gamma}W(t,1)\right]\left[\log(y)-\gamma\right]\frac{\mathrm{d}y}{y}+0 \\
= \quad & -\frac{\gamma^2}{t}\int_0^1 \left[W(t,u)/u - W(t,1)\right]\left[\log(u)+1\right]\mathrm{d}u \\
= \quad & -\frac{\gamma^2}{t}\int_0^1 W(t,u)/u\left[\log(u)+1\right]\mathrm{d}u
\end{aligned}$$

Use $\widehat{\gamma}_H(0,t) = \gamma + o_{\mathrm{P}}(1)$ uniformly in $t \in [t_0, 1-t_0]$ (from Example 2.4) and calculate covariances to obtain

$$t\sqrt{k}\left(\widehat{\gamma}_{MR}(0,t)-\gamma\right) \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \sigma_{\widehat{\gamma}_{MR},\gamma}W(t) \quad \text{in } D[t_0,1-t_0],$$

where $W(\cdot)$ is a standard Brownian motion and

$$\sigma^2_{\widehat{\gamma}_{MR},\gamma} = \gamma^2\int_0^1\int_0^1 \frac{r(x,y)}{xy}\left[\log(x)+1\right]\left[\log(y)+1\right]\mathrm{d}x\mathrm{d}y \overset{(2.7)}{=} 2\gamma^2\int_0^1 \frac{r(x,1)}{x}\mathrm{d}x. \tag{2.21}$$

Again, joint convergence as in (2.16) can be obtained by virtue of the joint convergence in (2.15).

Clearly, we have to consistently estimate $\sigma^2_{\widehat{\gamma},\gamma}$, the asymptotic variance of $\widehat{\gamma}(0,1)$. To that end we propose the following method. The basic idea is as follows: With only one sample $X_1,\ldots,X_n$ we can only estimate γ once with $\widehat{\gamma}(0,1)$ and infer nothing on the variance of the estimate. To get more estimates we calculate $\widehat{\gamma}(0,t)$ for $t \in (0,1]$. Calculating suitably normalized sample variances of all these estimates yields a consistent estimate of the variance of $\widehat{\gamma}(0,1)$, as shown in

Theorem 2.2. *Let $W(\cdot)$ denote a standard Brownian motion.*

(a) If for any $t_0 > 0$

$$t\sqrt{k}\left(\widehat{\gamma}\left(0,t\right)-\gamma\right) \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \sigma_{\widehat{\gamma},\gamma} W(t) \quad \text{in } D\left[t_0,1\right], \tag{2.22}$$

then for all sequences $t_n \downarrow 0$ tending to 0 not too fast,

$$\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{nor}} := \frac{1}{\log\left(n/(\lfloor nt_n\rfloor+1)\right)} \frac{k}{n} \sum_{i=\lfloor nt_n\rfloor+1}^{n} \left[\widehat{\gamma}\left(0,\frac{i}{n}\right)-\widehat{\gamma}\left(0,1\right)\right]^2 \underset{(n\to\infty)}{\xrightarrow{\mathrm{P}}} \sigma^2_{\widehat{\gamma},\gamma}.$$

(b) If for any $t_0 > 0$

$$(1-t)\sqrt{k}\left(\widehat{\gamma}\left(t,1\right)-\gamma\right) \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \sigma_{\widehat{\gamma},\gamma}\left(W(1)-W(t)\right) \quad \text{in } D\left[0,1-t_0\right], \tag{2.23}$$

then for all sequences $t_n \downarrow 0$ tending to 0 not too fast,

$$\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}} := \frac{1}{\log\left(n/(\lfloor nt_n\rfloor+1)\right)} \frac{k}{n} \sum_{i=0}^{n-(\lfloor nt_n\rfloor+2)} \left[\widehat{\gamma}\left(\frac{i}{n},1\right)-\widehat{\gamma}\left(0,1\right)\right]^2 \underset{(n\to\infty)}{\xrightarrow{\mathrm{P}}} \sigma^2_{\widehat{\gamma},\gamma}.$$

Remark 2.4. (a) See the proof of Theorem 2.2 below for why t_n must not approach 0 too fast.

(b) If **(A1)-(A4)** hold for $y_0 = 1$, then the convergences in (2.22) and (2.23) hold for any $t_0 > 0$ for the estimators given in Examples 2.3-2.5.

(c) In simulations we choose t_n as small as possible such that $\widehat{\gamma}\Big(0, \frac{\lfloor nt_n\rfloor+1}{n}\Big)$ (or $\widehat{\gamma}\Big(\frac{n-(\lfloor nt_n\rfloor+2)}{n}, 1\Big)$) is still well-defined for all choices of k. In fact, unreported simulations show that the estimates are quite robust with respect to the choice of t_n.

(d) Note that in the case of (e.g.) (a) in the above theorem

$$\frac{1}{\sigma^2_{\widehat{\gamma},\gamma}} \frac{k}{n} \sum_{i=\lfloor nt_n\rfloor+1}^{n} \left[\widehat{\gamma}\left(0,\frac{i}{n}\right)-\widehat{\gamma}\left(0,1\right)\right]^2 \overset{\mathcal{D}}{\approx} \frac{1}{n} \sum_{i=\lfloor nt_n\rfloor+1}^{n} \left[\frac{W(i/n)}{(i/n)} - W(1)\right]^2,$$

where the expectation of the right-hand side is (approximately)

$$\left[\left(\sum_{i=\lfloor nt_n\rfloor+1}^{n}\frac{1}{i}\right)-(1-t_n)\right]\sim\log\left(n/(\lfloor nt_n\rfloor+1)\right). \tag{2.24}$$

Hence, we use the left-hand side of (2.24) as a finite-sample correction for $\log\left(n/(\lfloor nt_n\rfloor+1)\right)$ in the estimator from Theorem 2.2 (a). A similar argument reveals that the finite-sample correction is also sensible for $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}}$.

(e) The result of the above theorem, which may be of independent interest, could be adapted to a wide range of estimators investigated in a change-point context, where limit results as (2.22) with $\sqrt{k}$ replaced by some other sequence tending to infinity and $\widehat{\gamma}(0,t)$ replaced by some other estimator (based on observations $X_1,\ldots,X_{\lfloor nt\rfloor}$) of an unknown parameter.

The joint convergences (as in (2.16)) established in the above examples and the continuous mapping theorem now allow us to easily derive the null distributions of our test statistics from (2.11).

Corollary 2.2. *If $\widehat{\sigma}^2_{\widehat{\gamma},\gamma}\in\{\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{nor}},\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}}\}$, then for the estimators from Examples 2.3-2.5 the convergences in* (2.12) *hold under the conditions of Corollary 2.1 with $y_0=1$.*

2.2.3 Results under the alternative

We will explore the behavior of our tests under two specific one-break alternatives:

$$\mathcal{H}_1^{\lessgtr}:\quad \gamma_1=\ldots=\gamma_{\lfloor nt^*\rfloor}\lessgtr\gamma_{\lfloor nt^*\rfloor+1}=\ldots=\gamma_n, \tag{2.25}$$

for some $t^*\in(t_0,1-t_0)$ and $\gamma_i>0$ for all $i=1,\ldots,n$. To avoid repetition in the following theorem we state conditions that must hold under $\mathcal{H}_1^{<}$ and that differ from the ones under $\mathcal{H}_1^{>}$ in parentheses (e.g., (2.27) below).

Theorem 2.3. *Under $\mathcal{H}_1^{>}$ $\left(\mathcal{H}_1^{<}\right)$ let the triangular array $\{X_{i,n}\}_{i=1,\ldots,n;n\in\mathbb{N}}$ be given by*

$$X_{i,n}:=\begin{cases}Y_i^{\mathrm{pre}}, & i\in I_{\mathrm{pre}}:=\{1,\ldots,\lfloor nt^*\rfloor\},\\ Y_i^{\mathrm{post}}, & i\in I_{\mathrm{post}}:=\{\lfloor nt^*\rfloor+1,\ldots,n\},\end{cases}$$

where $\left\{Y_i^{\mathrm{pre}}\right\}_{i\in\mathbb{N}}$ and $\left\{Y_i^{\mathrm{post}}\right\}_{i\in\mathbb{N}}$ both satisfy conditions **(A1)**-**(A4)** *with*

$$k_{\text{pre}},\ \gamma_{\text{pre}},\ U_{\text{pre}}(\cdot),\ r_{\text{pre}}(\cdot,\cdot),\ y_{0,\text{pre}} = \frac{1-t_0}{t_0}\quad \left(y_{0,\text{pre}} = \left(\frac{1-t_0}{t_0}\right)^{\gamma_{\text{post}}/\gamma_{\text{pre}}}\right) \qquad \textit{and}$$

$$k_{\text{post}},\ \gamma_{\text{post}},\ U_{\text{post}}(\cdot),\ r_{\text{post}}(\cdot,\cdot),\ y_{0,\text{post}} = \left(\frac{1-t_0}{t_0}\right)^{\gamma_{\text{pre}}/\gamma_{\text{post}}} \quad \left(y_{0,\text{post}} = \frac{1-t_0}{t_0}\right)$$

respectively. Suppose further that $q(\cdot)$ *satisfies* (2.13)*, and*

$$k_{\text{post}} = \mathcal{O}\left(k_{\text{pre}}\right),\ \textit{s.t. } k_{\text{post}} \left(\frac{U_{\text{post}}\left(\frac{n}{k_{\text{post}}}\right)}{U_{\text{pre}}\left(\frac{n}{k_{\text{pre}}}\right)}\right)^{1/\gamma_{\text{post}}} \underset{(n\to\infty)}{\longrightarrow} 0 \tag{2.26}$$

$$\left(k_{\text{pre}} = \mathcal{O}\left(k_{\text{post}}\right),\ \textit{s.t. } k_{\text{pre}} \left(\frac{U_{\text{pre}}\left(\frac{n}{k_{\text{pre}}}\right)}{U_{\text{post}}\left(\frac{n}{k_{\text{post}}}\right)}\right)^{1/\gamma_{\text{pre}}} \underset{(n\to\infty)}{\longrightarrow} 0\right). \tag{2.27}$$

Then, for the estimators from Examples 2.3-2.5, under $\mathcal{H}_1^{>}$

$$\sqrt{k_{\text{pre}}}\left(\widehat{\gamma}(0,t) - \gamma_{\text{pre}}\right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} B_{\text{pre}}(t)/t \qquad \textit{in } D\left[t_0, 1\right], \tag{2.28}$$

where

$$B_{\text{pre}}(t) := \begin{cases} \gamma_{\text{pre}} \int_0^1 W_{\text{pre}}\left(t_{\min}, u\frac{t}{t_{\min}}\right) J(u) \frac{\mathrm{d}u}{u}, & \textit{for } \widehat{\gamma}_{WLS}, \\ \gamma_{\text{pre}} \int_0^1 \left[W_{\text{pre}}\left(t_{\min}, u\frac{t}{t_{\min}}\right) - uW_{\text{pre}}\left(t_{\min}, \frac{t}{t_{\min}}\right)\right] \frac{\mathrm{d}u}{u}, & \textit{for } \widehat{\gamma}_{H}, \\ -\gamma_{\text{pre}} \int_0^1 W_{\text{pre}}\left(t_{\min}, u\frac{t}{t_{\min}}\right) \left[\log(u) + 1\right] \frac{\mathrm{d}u}{u}, & \textit{for } \widehat{\gamma}_{MR}, \end{cases}$$

with $t_{\min} := \min(t, t^*)$ *and* $W_{\text{pre}}(\cdot,\cdot)$ *as in Theorem 2.1 with* $r(\cdot,\cdot)$ *replaced by* $r_{\text{pre}}(\cdot,\cdot)$*, and under* $\mathcal{H}_1^{<}$

$$\sqrt{k_{\text{post}}}\left(\widehat{\gamma}(t,1) - \gamma_{\text{post}}\right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} B_{\text{post}}(t)/(1-t) \qquad \textit{in } D\left[t_0, 1\right],$$

where

$$B_{\text{post}}(t) := \begin{cases} \gamma_{\text{post}} \int_0^1 \widetilde{W}_{\text{post}}(t,u) J(u) \frac{\mathrm{d}u}{u}, & \textit{for } \widehat{\gamma}_{WLS}, \\ \gamma_{\text{post}} \int_0^1 \left[\widetilde{W}_{\text{post}}(t,u) - u\widetilde{W}_{\text{post}}(t,1)\right] \frac{\mathrm{d}u}{u}, & \textit{for } \widehat{\gamma}_{H}, \\ -\gamma_{\text{post}} \int_0^1 \widetilde{W}_{\text{post}}(t,u) \left[\log(u) + 1\right] \frac{\mathrm{d}u}{u}, & \textit{for } \widehat{\gamma}_{MR}, \end{cases}$$

with $\widetilde{W}_{\text{post}}(t,u) := W_{\text{post}}\left(1, u\frac{1-t}{1-t_{\max}}\right) - W_{\text{post}}\left(t_{\max}, u\frac{1-t}{1-t_{\max}}\right)$, $t_{\max} := \max(t, t^*)$*, and* $W_{\text{post}}(\cdot,\cdot)$ *as in Theorem 2.1 with* $r(\cdot,\cdot)$ *replaced by* $r_{\text{post}}(\cdot,\cdot)$*.*

Remark 2.5. (a) Quintos *et al.* (2001, Theorem 3) show that the Hill estimator applied to an i.i.d. sample with one break in the tail index as in (2.25) converges in probability to $\max(\gamma_{\text{pre}}, \gamma_{\text{post}})$. Theorem 2.3 obviously substantially generalizes this result.

(b) Theorem 2.3 does not make any assumption on the dependence between Y_i^{pre} and Y_i^{post}.

(c) For the time series models from Examples 2.1 and 2.2 conditions **(A1)-(A3)** were satisfied for sequences k with lower and upper bound

$$\log^2(n)\log^4(\log n) = o(k) \qquad \text{and} \qquad k = o(n^{\xi}), \quad \xi \in (0,1)$$

(recall (2.2) and (2.8)). Hence for a sample with a break in the tail index it does not seem to be overly restrictive to assume $k = k_{\text{pre}} = k_{\text{post}}$, which is what we do in the following.

(d) Under $\mathcal{H}_1^{>}$ ($\mathcal{H}_1^{<}$) condition (2.26) ((2.27)) ensures that the part of the sequential tail empirical process appertaining to the post- (pre-) break period is asymptotically negligible (see the proof of Theorem 2.3 below).

How stringent is (2.26)? (A similar argument also holds for (2.27).) For any $\varepsilon > 0$ and n sufficiently large, note that

$$\begin{aligned}
\frac{U_{\text{post}}\left(\frac{n}{k_{\text{post}}}\right)}{U_{\text{pre}}\left(\frac{n}{k_{\text{pre}}}\right)} &= \frac{\left(\frac{n}{k_{\text{post}}}\right)^{\gamma_{\text{post}}+\varepsilon}\left(\frac{n}{k_{\text{post}}}\right)^{-\varepsilon} L_{\text{post}}\left(\frac{n}{k_{\text{post}}}\right)}{\left(\frac{n}{k_{\text{pre}}}\right)^{\gamma_{\text{pre}}-\varepsilon}\left(\frac{n}{k_{\text{pre}}}\right)^{\varepsilon} L_{\text{pre}}\left(\frac{n}{k_{\text{pre}}}\right)} \\
&< \left(\frac{n}{k_{\text{post}}}\right)^{\gamma_{\text{post}}+\varepsilon}\left(\frac{n}{k_{\text{pre}}}\right)^{-\gamma_{\text{pre}}+\varepsilon},
\end{aligned}$$

where $L_{\text{pre(post)}}(x) = x^{-\gamma_{\text{pre(post)}}}U_{\text{pre(post)}}(x) \in RV_0$ and Bingham, Goldie and Teugels' (1987) Proposition 1.3.6 (v) was used for the inequality. If $k = k_{\text{pre}} = k_{\text{post}} = n^{\xi}$ for some $\xi \in (0,1)$, then (2.26) is satisfied for $\xi < 1 - \gamma_{\text{post}}/\gamma_{\text{pre}}$. That is, for small breaks, i.e., $\gamma_{\text{pre}} - \gamma_{\text{post}}$ close to 0, k must be rather small relative to n. Similar arguments apply to (2.27).

(In-)Consistency results can now easily be proved:

Corollary 2.3. *Under the conditions of Theorem 2.3 with $k = k_{\text{pre}} = k_{\text{post}}$ we have for the estimators from Examples 2.3-2.5 and for all sequences $t_n \downarrow 0$ tending to zero not too fast:*

(a) $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{nor}} \xrightarrow[(n\to\infty)]{\text{P}} \sigma^2_{\widehat{\gamma},\gamma_{\text{pre}}}$ *under* $\mathcal{H}_1^{>}$*,*

(b) $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}} \underset{(n\to\infty)}{\overset{\mathrm{P}}{\longrightarrow}} \sigma^2_{\widehat{\gamma},\gamma_{\mathrm{post}}}$ *under* $\mathcal{H}_1^{<}$.

Using $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{nor}}$ *or* $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}}$ *according as* $\mathcal{H}_1^{>}$ *or* $\mathcal{H}_1^{<}$*, we further have:*

(c) The tests based on Q_{seq} *and* Q_{rol} *are consistent under* $\mathcal{H}_1^{\lessgtr}$*, where for* Q_{rol} *the additional assumption* $t^* \in (t_0, 1-2t_0)$ $\big(t^* \in (2t_0, 1-t_0)\big)$ *has to hold under* $\mathcal{H}_1^{>}$ $\left(\mathcal{H}_1^{<}\right)$.

(d) The test based on Q_{rec} $(Q^{\leftarrow}_{\mathrm{rec}})$ *is consistent under* $\mathcal{H}_1^{<}$ $\left(\mathcal{H}_1^{>}\right)$*, whereas under* $\mathcal{H}_1^{>}$ $\left(\mathcal{H}_1^{<}\right)$ *we have*

$$Q^{(\leftarrow)}_{\mathrm{rec}} \underset{(n\to\infty)}{=} \mathcal{O}_{\mathrm{P}}(1).$$

Remark 2.6. For an estimator of the change point t^* that is consistent under weak conditions we refer to Kim and Lee (2009, Theorem 3).

2.3 Simulations

This section investigates the finite-sample properties of our tests for specific models from Examples 2.1 and 2.2. We do so only for Q_{rec} and $Q^{\leftarrow}_{\mathrm{rec}}$, because we want to explore the differences between the tail index estimators and not between the different test statistics in (2.11). The latter has already been done in the literature (Quintos *et al.*, 2001; Kim and Lee, 2011). We just remark that the qualitative conclusions from the other studies hold here as well. The $Q^{(\leftarrow)}_{\mathrm{rec}}$- and Q_{seq}-test have slightly better size than the Q_{rol}-test, presumably because the estimates $\widehat{\gamma}(t, t+t_0)$ are always based on relatively small rolling windows. Under the alternative, when the tests based on $Q^{(\leftarrow)}_{\mathrm{rec}}$ are consistent, they have the highest power of all alternatives, which is why we focus on these tests here.

We use $t_0 = 0.2$, sample sizes of $n = 500$ and $n = 2000$, and $t_n = 50/n$. With this choice of t_n all estimates $\widehat{\gamma}\Big(0, \frac{\lfloor nt_n\rfloor+1}{n}\Big)$ and $\widehat{\gamma}\Big(\frac{n-(\lfloor nt_n\rfloor+2)}{n}, 1\Big)$ remained well-defined and t_n is reasonably small, as required by Theorem 2.2. Table 2.1 shows critical values, obtained by 100,000 realizations of the approximations to the limit distribution $\sup_{t\in[t_0,1-t_0]} \{W(t) - tW(1)\}^2$ from (2.12) (where the Brownian motion itself was generated from 100,000 independent normally distributed r.v.s). We use the the estimators $\widehat{\gamma}_H$, $\widehat{\gamma}_{MR}$ and $\widehat{\gamma}_{CV_\theta}$ for $\theta = 1$. The corresponding tests will be denoted H, MR and CV_1.

In order for tests to be consistent, we estimate $\sigma^2_{\widehat{\gamma},\gamma}$ using $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{nor}}$ for $Q^{\leftarrow}_{\mathrm{rec}}$ (under $\mathcal{H}_0$ and $\mathcal{H}_1^{>}$) and $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}}$ for Q_{rec} (under $\mathcal{H}_1^{<}$). We modify $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{nor}}$ and $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}}$ by requiring that they be at least as large as the (consistent) variance estimate in the independent case, i.e., $\widehat{\gamma}_H^2(0,1)$ for $\widehat{\gamma}_H(0,1)$, $2\widehat{\gamma}_{MR}^2(0,1)$ for $\widehat{\gamma}_{MR}(0,1)$ and $\frac{2\theta+2}{2\theta+1}\widehat{\gamma}_{CV_\theta}^2(0,1)$ for $\widehat{\gamma}_{CV_\theta}(0,1)$. This is warranted by the observation that our models from Examples 2.1 and 2.2 satisfy the conditions of Drees (2003, Prop. 2.1), whence $r(x,y) \geq \min(x,y)$. (Note that $r(x,y) = \min(x,y)$ under independence and **(A4)**.) Hence, the asymptotic variances given in (2.18), (2.19) and (2.21) cannot be lower under dependence than under independence, such that our modified estimators are still consistent for $\sigma^2_{\widehat{\gamma},\gamma}$ under independence *and* dependence.

α_q	0.50	0.60	0.70	0.80	0.90	0.95	0.99
α_q-quantile	0.650	0.767	0.918	1.128	1.478	1.821	2.653

Table 2.1: Quantiles of $\sup_{t\in[t_0,1-t_0]} \left\{W(t) - tW(1)\right\}^2$ for $t_0 = 0.2$

Concretely, we simulate from the two data generating processes (DGPs)

$$X_i = 0.5 \cdot X_{i-1} + Z_i, \tag{AR}$$

$$X_i^2 = \left(\alpha_0 + \alpha_1 \cdot X_{i-1}^2\right) Z_i^2. \tag{ARCH}$$

Remark 2.7. (a) In the AR(1) case it is also possible to use the change point test proposed in Kim and Lee (2012), which is based on AR(p)-residuals. However, in the context of extreme quantile estimation Drees (2008, Section 2) cautions against using residual-based tail index estimators for AR(p)-models, since they can be very sensitive to ever so slight misspecifications. We therefore advocate using non model-based estimators in a change point context as well.

(b) For the (G)ARCH-model with innovations that have finite $(4+\delta)$-th moments, there exist more precise estimators of the tail index (e.g., in Berkes, Horváth and Kokoszka, 2003, and Chan *et al.*, 2013) in the sense of being $\sqrt{n}$-consistent instead of the slower $\sqrt{k}$. Using these estimators in a change point test could potentially result in more powerful tests. However, as in part (a) of this remark, slight departures from the model could then lead to severe misestimation of tail parameters.

For the ARCH-model one often uses $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ or the normalized $\sqrt{\frac{\nu-2}{\nu}}t_\nu$ ($\nu > 2$) with unit variance. For standardized t_ν-innovations the first moment condition in (2.9) implies that $\alpha_1 \in (0, \exp\{\psi(\nu/2) - \psi(1/2)\} / \{\nu - 2\})$, where $\psi(z) = \Gamma'(z)/\Gamma(z)$ denotes the digamma function, since

$$0 > \mathbb{E}\log(\alpha_1 Z_i^2) = \log(\alpha_1) + \log\left(\frac{\nu-2}{\nu}\right) + \log(\nu) + \psi(1/2) - \psi(\nu/2), \tag{2.29}$$

Test	Size	DGP	$n = 500$				$n = 2000$			
			k/n							
			0.08	0.12	0.16	0.2	0.08	0.12	0.16	0.2
H	0.05	(AR)	4.5	4.8	4.2	4.1	5.6	4.8	4.5	4.1
	0.01		1.3	1.0	1.1	0.8	1.2	0.9	0.8	0.7
MR	0.05		2.9	2.5	2.0	1.5	3.4	3.2	2.7	2.3
	0.01		1.1	0.7	0.4	0.3	0.8	0.8	0.6	0.4
CV_1	0.05		2.0	1.9	1.9	1.7	3.8	3.8	3.4	3.2
	0.01		0.4	0.3	0.2	0.3	0.7	0.6	0.5	0.3
H	0.05	(ARCH)	5.7	5.5	4.4	4.4	6.9	6.1	5.4	5.2
	0.01		1.6	1.0	1.0	0.9	1.8	1.1	1.0	1.1
MR	0.05		3.0	2.8	2.4	1.9	4.6	3.8	3.2	3.0
	0.01		1.1	1.0	0.7	0.6	1.6	1.1	0.7	0.7
CV_1	0.05		2.8	2.1	1.8	1.5	4.5	4.4	4.6	4.4
	0.01		0.8	0.6	0.4	0.2	0.9	0.7	0.7	0.7

Table 2.2: Empirical sizes of $Q^{\leftarrow}_{\text{rec}}$-tests in % for n realizations of (AR) and (ARCH)

where $\log(t_\nu^2)/2 \sim \log(F_{1,\nu})/2$ follows Fisher's z-distribution with mean

$$\left[\log(\nu) + \psi(1/2) - \psi(\nu/2)\right]/2.$$

Remark 2.8. The only existing change point test known to the author for ARCH data in Quintos *et al.* (2001) relies on standard-normally distributed innovations, whereas our tests permit, e.g., $\sqrt{\frac{\nu-2}{\nu}}t_\nu$-distributed innovations for $\nu > 2$. If only standard-normally distributed innovations are permitted, tail index break tests degenerate to tests for parameter constancy for α_1 (recall from Example 2.2 that the tail index of an ARCH(1) can only be changed by varying α_1 or the distribution of Z_1). We venture to claim that tests for parameter constancy for GARCH(p,q) models as proposed in Berkes *et al.* (2004) then perform better, as more observations are effectively used in the estimation of α_1.

For the results under the null in Table 2.2 we choose $Z_i \sim \sqrt{\frac{\nu-2}{\nu}}t_\nu$ with $\nu = 5/2$ for the ARCH(1)-model along with $\alpha_0 = 0.01$ and $\alpha_1 = 0.95$, i.e., tail index equal to 1.01 determined from $\mathbb{E}(\alpha_1 Z_i^2)^\alpha = 1$. Note that by choosing $\nu = 5/2$ the innovations barely have existing second moments, which is required in ARCH-type models. Note further that by (2.29) the choice of $\nu = 5/2$ necessitates $\alpha_1 \in (0, 11.34...)$, which is of course satisfied for our particular choice of α_1. For the AR(1)-model we also use $Z_i \sim t_\nu$ with $\nu = 5/2$, i.e., tail index equal to 2.5. Hence, the process in (AR) does have finite second moments, while that in (ARCH) does not.

For both models the results show that, by and large, sizes only slightly decrease in k. This is encouraging since the choice of k can be a very sensitive issue in tail index estimation, see, e.g., Section 4.4.2 in Resnick (2007) for a Hill horror plot and some references on the topic. As a referee pointed out, this may be explained by the canceling of bias terms (that arise in tail index estimation for large k) in (2.11). For $n = 500$ most tests are conservative for both models. The convergence to the nominal level for $n = 2000$ is satisfactory for the H and the CV_1 test for a wide range of k's, while the MR test is still somewhat conservative.

To examine power we simulate according to model (AR) again, only that now $Z_i := Z_{i,n}$ with

$$Z_{i,n} \sim \begin{cases} t_2, & i \leq \lfloor nt^* \rfloor, \\ t_\nu, & i > \lfloor nt^* \rfloor, \end{cases} \tag{2.30}$$

where we choose $t^* = 0.25, 0.5$ and $\nu \in [2, 4]$. Hence, there is a break in the tail index of $\{X_i\}$ from 2 to ν, i.e., lighter post-break tails. To investigate power for a nonlinear model as well, we simulate from

$$X_{i,n}^2 := \begin{cases} (0.01 + 0.45 \cdot X_{i-1,n}^2) Z_i^2, & i \leq \lfloor nt^* \rfloor, \\ (0.01 + \alpha_1 \cdot X_{i-1,n}^2) Z_i^2, & i > \lfloor nt^* \rfloor, \end{cases} \tag{2.31}$$

where again $t^* = 0.25, 0.5$, $\alpha_1 \in [0.45, 0.95]$ and $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. This ARCH(1)-model with a break in α_1 has tails varying from thinner ($\alpha = 2.67$ corresponding to $\alpha_1 = 0.45$) to thicker ($\alpha = 1.07$ corresponding to $\alpha_1 = 0.95$). Throughout we take $n = 2000$.

The simulation results for different values of ν in (2.30) and α_1 in (2.31) are displayed in Figure 2.1. We choose $k = 0.2 \cdot n$ for both models, because for these values of k the differences in size (i.e., $\nu = 2$ in the AR(1)-case and $\alpha_1 = 0.45$ in the ARCH(1)-case) are smallest, such that direct power comparisons are more meaningful. Figure 2.1 (a) displays the results for the AR(1)-model with innovations as in (2.30) using the $Q_{\text{rec}}^{\leftarrow}$-test, which is consistent. In the bottom part, where results are shown for $t^* = 0.5$, tests have roughly comparable properties. It is notable that, despite being slightly conservative, the MR test offers higher power than the other two tests, the more so the larger ν. The difference for $\nu = 4$ is between 13 and 15 percentage points. This is even more apparent when $t^* = 0.25$, where the MR test performs only marginally worse than before, but the CV_1 test and the H test in particular lose sizable amounts of power. Here the biggest difference is as high as 38 percentage points.

Panel (b) shows results for the ARCH(1) with a break in α_1 using Q_{rec}. The top part for $t^* = 0.25$ shows that the MR test has comparable power as the other

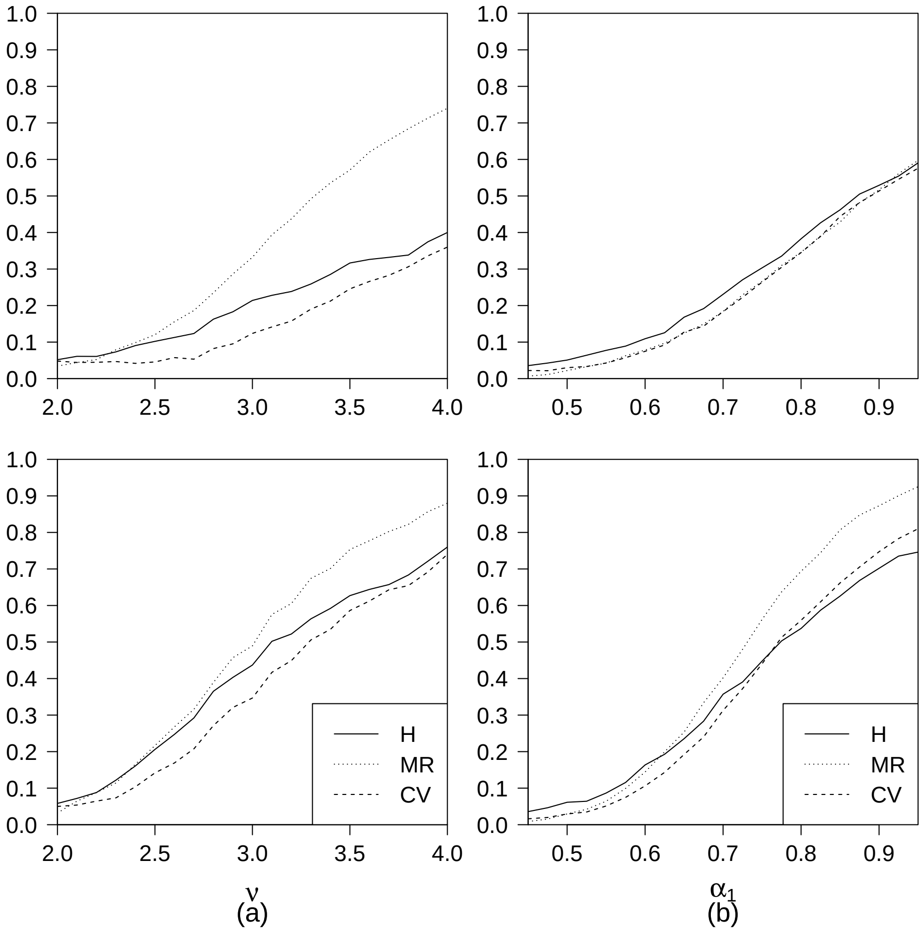

Figure 2.1: (a) Fraction of rejections for $Q_{\text{rec}}^{\leftarrow}$-test using (AR) with innovations (2.30) and (b) for Q_{rec}-test using (2.31) with $t^* = 0.25$ (top), $t^* = 0.5$ (bottom)

two tests despite being more conservative. When $t^* = 0.5$ (bottom part) it seems to gain more power than the other two offerings, so that power for the MR test is 18 percentage points higher than that for the H test for $\alpha_1 = 0.95$. In light of the simulation evidence in Wagner and Marsh (2004) already mentioned in the motivation, this was to be expected.

In the upper left plot in Figure 2.1 the CV_1 test seems to suffers slightly from nonmonotonic power, a well-known phenomenon in the literature on change point detection (Vogelsang, 1997, 1999), i.e., it does not show increasing power in distance from the null in some ranges. In the context of mean-shift detection Vogelsang (1999) identifies long-run variance estimates as one major source of nonmonotonicity. We also find indications for this here. For $\nu = 2$ the average estimate of $\widehat{\sigma}^2_{\widehat{\gamma}_{CV_1},\gamma}$ $(\widehat{\sigma}^2_{\widehat{\gamma}_{MR},\gamma})$ over all 5000 replications is 0.74 (0.82), while for $\nu = 2.7$ it is 0.95 (0.70). Hence, while for the MR test the denominator of $Q^{\leftarrow}_{\text{rec}}$ even decreased, it increased markedly for the CV_1 test. Apparently, it increased roughly proportionately to the numerator of $Q^{\leftarrow}_{\text{rec}}$ for CV_1, which could be a reason for the flat profile for $\nu \in [2, 2.7]$.

All in all, our simulations reveal reasonable size of our tests for quite a wide range of k's with some conservative tendencies. Under the alternative we find the MR test to clearly deliver the best results, with the H and CV_1 test performing similarly.

2.4 Proofs

In the following K, K_1, K_2 and $\widetilde{\delta}$ denote large and small positive constants that may change from line to line. $D[t_0, 1]$ denotes the space of càdlàg functions equipped with the Skorohod metric and the Borel σ-field $\mathcal{D}[t_0, 1]$. For brevity put $D^2 := D([t_0, 1] \times [0, y_0 + \delta])$ $(\delta \geq 0)$ for the space of two-parameter càdlàg functions on $[t_0, 1] \times [0, y_0 + \delta]$, which is equipped with the multiparameter extension of the Skorohod metric (cf. Bickel and Wichura, 1971, p. 1662) and the Borel σ-field $\mathcal{D}^2$. As usual, define $\|\cdot\|_2$ to be the Euclidean metric, $|\cdot|$ applied to a set the cardinality and $\sum_i^j := 0$ for $i > j$.

To derive weighted weak convergence results involving the weight function $q(\cdot)$ we may assume w.l.o.g., as in the proof of Drees (2000, Theorem 2.2), that for some $\vartheta > 0$ sufficiently small

$$q(y) = y^{\nu} |\log y|^{\mu}, \ y \in (0, \vartheta], \quad \text{s.t. } q \text{ is increasing and } q/Id \text{ decreasing on } (0, \vartheta], \tag{2.32}$$

where $Id(\cdot)$ denotes the identity function.

In a first step we will consider uniformly distributed r.v.s $U_i \sim \mathcal{U}[0, 1]$ and then suitably apply this result to X_i satisfying **(A4)**. To this end we need the following analogs of conditions **(A1)**-**(A3)**:

(U1) $\{U_i\}_{i \in \mathbb{N}}$ is a strictly stationary β-mixing process with mixing coefficients $\beta(\cdot)$, such that
$$\lim_{n \to \infty} \frac{n}{r_n} \beta(l_n) + \frac{r_n}{\sqrt{k}} \log^2(k) = 0$$
for sequences $\{l_n\}_{n \in \mathbb{N}} \subset \mathbb{N}, \{r_n\}_{n \in \mathbb{N}} \subset \mathbb{N}$ tending to infinity with $l_n = o(r_n)$, $r_n = o(n)$.

(U2) There exists a function $r(x,y)$, such that for all $x, y \in [0, y_0+\delta]$

$$\lim_{n\to\infty} \frac{n}{r_n k} \operatorname{Cov}\left(\sum_{i=1}^{r_n} I_{\left\{U_i > 1-\frac{k}{n}x\right\}}, \sum_{j=1}^{r_n} I_{\left\{U_j > 1-\frac{k}{n}y\right\}} \right) = r(x,y).$$

(U3) For some constant $C > 0$

$$\frac{n}{r_n k} \mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{1-\frac{k}{n}y < U_i \le 1-\frac{k}{n}x\right\}} \right]^4 \le C(y-x) \qquad \forall\, 0 \le x < y \le y_0+\delta,\ n \in \mathbb{N}.$$

We start with a non-weighted weak convergence result for the sequential tail empirical process for the $\{U_i\}$:

Theorem 2.4. *Suppose* **(U1)**-**(U3)** *hold. Then*

$$\sqrt{k}\left\{ \frac{1}{k} \sum_{i=1}^{\lfloor nt \rfloor} \left[I_{\left\{U_i > 1-\frac{k}{n}y\right\}} - \frac{k}{n}y \right] \right\} \xrightarrow[(n\to\infty)]{\mathcal{D}} W(t,y) \quad \text{in } D^2, \tag{2.33}$$

where $\{W(t,y)\}$ *is a continuous zero-mean Gaussian process with covariance function*

$$\operatorname{Cov}\left(W(t_1,y_1), W(t_2,y_2)\right) = \min(t_1,t_2)\, r(y_1,y_2). \tag{2.34}$$

Proof. For notational convenience and w.l.o.g. $\delta = 0$. We use a classical 'big block - small block' approach, where the small blocks are asymptotically negligible. For $t \in [0,1]$ define

$$m_n(t) := \left\lfloor \frac{\lfloor nt \rfloor}{r_n + l_n} \right\rfloor$$

and for $j = 1, \ldots, m_n(1)$ define I_j (the big blocks) and J_j (the small blocks) to be consecutive blocks of integers of length $|I_j| = r_n$ and $|J_j| = l_n$, i.e.,

$$I_1 = \{1, \ldots, r_n\},\ J_1 = \{r_n+1, \ldots, r_n+l_n\},\ \text{etc.}$$

Choose the length of $I_{m_n(t)+1}$ such that the integers $\{1, \ldots, \lfloor nt \rfloor\}$ are covered. Now decompose

$$\sqrt{k}\left\{ \frac{1}{k} \sum_{i=1}^{\lfloor nt \rfloor} \left[I_{\left\{U_i > 1-\frac{k}{n}y\right\}} - \frac{k}{n}y \right] \right\} = \sum_{j=1}^{m_n(t)} Y_j^I(y) + \sum_{j=1}^{m_n(t)} Y_j^J(y) + R_n(t,y),$$

where

$$Y_j^I(y) = \frac{1}{\sqrt{k}} \sum_{i \in I_j} \left[I_{\left\{U_i > 1 - \frac{k}{n} y\right\}} - \frac{k}{n} y \right],$$

$$Y_j^J(y) = \frac{1}{\sqrt{k}} \sum_{i \in J_j} \left[I_{\left\{U_i > 1 - \frac{k}{n} y\right\}} - \frac{k}{n} y \right],$$

$$R_n(t, y) = \frac{1}{\sqrt{k}} \sum_{i \in I_{m_n(t)+1}} \left[I_{\left\{U_i > 1 - \frac{k}{n} y\right\}} - \frac{k}{n} y \right].$$

We will consider these three terms separately. First, noting that the cardinality $\left|I_{m_n(t)+1}\right| \leq r_n + l_n - 1$,

$$0 \leq \sup_{(t,y) \in [t_0,1] \times [0,y_0]} \left|R_n(t, y)\right| \leq 2 \frac{r_n + l_n - 1}{\sqrt{k}} \underset{(n \to \infty)}{\overset{\textbf{(U1)}}{\longrightarrow}} 0. \tag{2.35}$$

Second, set $L_n(t, y) = \sum_{j=1}^{m_n(t)} Y_j^I(y)$ and define the measurable mapping

$$M_n : \left(D^{m_n}[0, y_0], \mathcal{D}^{m_n}[0, y_0]\right) \to \left(D^2, \mathcal{D}^2\right)$$
$$M_n\left(t, x_1(\cdot), \ldots, x_{m_n}(\cdot)\right) = \sum_{i=1}^{m_n(t)} x_i(y), \qquad (t, y) \in [t_0, 1] \times [0, y_0].$$

Then for $H \in \mathcal{D}^2$ using Lemma 2 of Eberlein (1984) and **(U1)**

$$\begin{aligned} \mathrm{P}(L_n \in H) &= \mathrm{P}\left(\left(Y_1^I(\cdot), \ldots, Y_{m_n}^I(\cdot)\right) \in M_n^{-1}(H)\right) \\ &= \widetilde{\mathrm{P}}\left(\left(\widetilde{Y}_1^I(\cdot), \ldots, \widetilde{Y}_{m_n}^I(\cdot)\right) \in M_n^{-1}(H)\right) + \mathcal{O}\left(m_n \beta(l_n)\right) \\ &\overset{\textbf{(U1)}}{=} \widetilde{\mathrm{P}}(\widetilde{L}_n \in H) + o(1), \end{aligned} \tag{2.36}$$

where $\widetilde{L}_n(t, y) = \sum_{j=1}^{m_n(t)} \widetilde{Y}_j^I(y)$ and the $\widetilde{Y}_j^I(\cdot)$ are i.i.d. copies of $Y_1^I(\cdot)$ defined on the product probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathrm{P}}) := \otimes_{i=1}^{\infty}(D[0, y_0], \mathcal{D}[0, y_0], \mathrm{P}_{Y_i})$ via

$$\tilde{Y}_i : (\tilde{\Omega}, \tilde{\mathcal{A}}) \to (D[0, y_0], \mathcal{D}[0, y_0]), \quad \tilde{Y}_i(\omega) := \pi_i(\omega) := \omega_i.$$

Now Corollary 3.3 of Hill (2009) implies

$$\widetilde{L}_n(t, y) \underset{(n \to \infty)}{\overset{\mathcal{D}}{\longrightarrow}} W(t, y) \quad \text{in } D^2, \tag{2.37}$$

where $\{W(t,y)\}$ is a zero-mean Gaussian process with continuous paths along t and y. For the covariance structure of the process consider weak convergence of the $\mathbb{R}^2$-valued random vector

$$\begin{pmatrix} \widetilde{L}_n(t_1,y_1) \\ \widetilde{L}_n(t_2,y_2) \end{pmatrix}$$

or, using the Cramér-Wold device, of

$$\begin{aligned} a\widetilde{L}_n(t_1,y_1) + b\widetilde{L}_n(t_2,y_2) &= b\left[\widetilde{L}_n(t_2,y_2) - \widetilde{L}_n(t_1,y_2)\right] + \left[a\widetilde{L}_n(t_1,y_1) + b\widetilde{L}_n(t_1,y_2)\right] \\ &=: A_n + B_n. \end{aligned}$$

for arbitrary $a, b \in \mathbb{R}$. Observe that A_n and B_n are independent for each n and hence it suffices to consider weak convergence of A_n and B_n separately. W.l.o.g. let $t_1 \leq t_2$.

For A_n we have

$$\widetilde{L}_n(t_2,y_2) - \widetilde{L}_n(t_1,y_2) = \sum_{j=m_n(t_1)+1}^{m_n(t_2)} \widetilde{Y}_j^I(y_2).$$

Then **(U2)** implies

$$\begin{aligned} s_n^2 := \sum_{j=m_n(t_1)+1}^{m_n(t_2)} \operatorname{Var}\left(\widetilde{Y}_j^I(y_2)\right) &= \frac{m_n(t_2) - m_n(t_1)}{k} \operatorname{Var}\left(\sum_{i=1}^{r_n} I_{\left\{U_i > 1 - \frac{k}{n} y_2\right\}}\right) \\ &\underset{(n\to\infty)}{\longrightarrow} (t_2 - t_1)\, r(y_2,y_2) =: \sigma_A^2. \end{aligned} \tag{2.38}$$

The Lyapunov condition (cf., e.g., Billingsley, 1968, Theorem 7.3) is satisfied (for $\delta = 2$), since

$$\frac{1}{s_n^4} \sum_{j=m_n(t_1)+1}^{m_n(t_2)} \mathbb{E}[\widetilde{Y}_j^I(y_2)]^4 \leq K \frac{1}{k} \frac{n}{r_n k} \mathbb{E}\left[\sum_{i=1}^{r_n}\left(I_{\left\{U_i > 1 - \frac{k}{n} y_2\right\}} - \frac{k}{n} y_2\right)\right]^4 \underset{(n\to\infty)}{\overset{\textbf{(U3)}}{\longrightarrow}} 0. \tag{2.39}$$

Using (2.38) we thus get

$$A_n \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}(0, b^2\sigma_A^2). \tag{2.40}$$

For B_n we have

$$a\widetilde{L}_n(t_1,y_1) + b\widetilde{L}_n(t_1,y_2) = \sum_{j=1}^{m_n(t_1)} \left(a\widetilde{Y}_j^I(y_1) + b\widetilde{Y}_j^I(y_2)\right).$$

Reasoning similarly as for the weak convergence of A_n (using Loève's c_r inequality for the analog of (2.39)) we get

$$B_n \xrightarrow[(n\to\infty)]{\mathcal{D}} \mathcal{N}(0, \sigma_B^2), \tag{2.41}$$

where $\sigma_B^2 = a^2 t_1 r(y_1, y_1) + 2ab t_1 r(y_1, y_2) + b^2 t_1 r(y_2, y_2)$. Combining (2.40) and (2.41) gives

$$A_n + B_n = a\widetilde{L}_n(t_1, y_1) + b\widetilde{L}_n(t_2, y_2) \xrightarrow[(n\to\infty)]{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = b^2\sigma_A^2 + \sigma_B^2 = a^2 t_1 r(y_1, y_1) + 2ab t_1 r(y_1, y_2) + b^2 t_2 r(y_2, y_2)$, i.e.,

$$\begin{pmatrix} \widetilde{L}_n(t_1, y_1) \\ \widetilde{L}_n(t_2, y_2) \end{pmatrix} \xrightarrow[(n\to\infty)]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \stackrel{\mathcal{D}}{=} \begin{pmatrix} W(t_1, y_1) \\ W(t_2, y_2) \end{pmatrix},$$

where

$$\mathbf{\Sigma} = \begin{pmatrix} t_1 r(y_1, y_1) & t_1 r(y_1, y_2) \\ t_1 r(y_1, y_2) & t_2 r(y_2, y_2) \end{pmatrix}.$$

Thus, $\{W(t, y)\}$ has the claimed covariance structure in (2.34). By (2.36) we also have

$$L_n(t, y) \xrightarrow[(n\to\infty)]{\mathcal{D}} W(t, y) \quad \text{in } D^2. \tag{2.42}$$

Third, in view of (2.35) and (2.42) it remains to prove

$$\sup_{\substack{t\in[t_0,1] \\ y\in[0,y_0]}} \left| \sum_{j=1}^{m_n(t)} Y_j^J(y) \right| \underset{(n\to\infty)}{=} o_{\mathrm{P}}(1). \tag{2.43}$$

Set

$$S_m(y) := \sum_{j=1}^{m} \widetilde{Y}_j^J(y) \qquad \text{and} \qquad \|S_m\| := \sup_{y\in[0,y_0]} \left|S_m(y)\right|,$$

where $\widetilde{Y}_j^J(\cdot)$ are i.i.d. copies of $Y_j^J(\cdot)$ as above. Similarly as in (2.36), to show that $\sum_{j=1}^{m_n(t)} Y_j^J(y)$ is asymptotically negligible, it suffices to do so for $\sum_{j=1}^{m_n(t)} \widetilde{Y}_j^J(y)$. To this end the Ottaviani inequality (cf., e.g., Shorack and Wellner, 1996, Proposition A.1.1) yields for any $\varepsilon > 0$

$$\mathrm{P}\left\{\sup_{\substack{t\in[t_0,1]\\ y\in[0,y_0]}}\left|\sum_{j=1}^{m_n(t)}\widetilde{Y}_j^J(y)\right| > 2\varepsilon\right\} \leq \mathrm{P}\left\{\max_{m\in\{1,\dots,m_n(1)\}}\|S_m\| > 2\varepsilon\right\}$$
$$\leq \frac{\mathrm{P}\left\{\|S_{m_n(1)}\| > \varepsilon\right\}}{1-\max_{m\in\{1,\dots,m_n(1)\}}\mathrm{P}\left\{\|S_m\| > \varepsilon\right\}}.$$

We show that $\mathrm{P}\left\{\|S_m\| > \varepsilon\right\} = o(1)$ uniformly in $m = 1, \dots, m_n(1)$. For this let $\Delta = \Delta_n > 0$ be a sequence, s.t. $\Delta = \mathcal{O}(k^{-1/2})$ and $y_0/\Delta \in \mathbb{N}$. Observe that (because of $m \leq n/r_n$) for all $y \in \big[(i-1)\Delta, i\Delta\big]$

$$S_m((i-1)\Delta) - \underbrace{\frac{l_n}{r_n}\sqrt{k}\Delta}_{\to 0} \leq S_m(y) \leq S_m(i\Delta) + \underbrace{\frac{l_n}{r_n}\sqrt{k}\Delta}_{\to 0},$$

from which we conclude via Markov's inequality

$$\mathrm{P}\left\{\|S_m\| > \varepsilon\right\} \leq P\left\{\max_{i\in\{0,\dots,y_0/\Delta\}}\left|S_m(i\Delta)\right| > \varepsilon/2\right\} \leq (\varepsilon/2)^{-4}\mathbb{E}\left[\max_{i\in\{0,\dots,y_0/\Delta\}}\left|S_m(i\Delta)\right|^4\right].$$

First we bound $\mathbb{E}\left[\left|S_m(i\Delta)\right|^4\right]$ by arguments similar to Rootzén (2009, p. 479). We have

$$\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U_i > 1-\frac{k}{n}y\right\}}\right]^4 \geq \mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U_i > 1-\frac{k}{n}y\right\}}\right]^2, \tag{2.44}$$

because the sum of the indicators is $\mathbb{N}_0$-valued, and (also using strict stationarity) for $p = 2, 4$

$$\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U_i > 1-\frac{k}{n}y\right\}}\right]^p \geq \mathbb{E}\left[\sum_{w=1}^{\lfloor r_n/l_n\rfloor}\sum_{i=(w-1)l_n+1}^{wl_n} I_{\left\{U_i > 1-\frac{k}{n}y\right\}}\right]^p \geq \left\lfloor\frac{r_n}{l_n}\right\rfloor \mathbb{E}\left[\sum_{i=1}^{l_n} I_{\left\{U_i > 1-\frac{k}{n}y\right\}}\right]^p,$$

whence $\mathbb{E}\left[\widetilde{Y}_j^J(y)\right]^p \leq K\frac{l_n}{n}k^{1-p/2}y$ with **(U3)**. Rosenthal's inequality now implies

$$\mathbb{E}\left[\left|S_m(i\Delta)\right|^4\right] \leq K\left\{m\mathbb{E}\left[\widetilde{Y}_j^J(i\Delta)\right]^4 + \left(m\mathbb{E}\left[\widetilde{Y}_j^J(i\Delta)\right]^2\right)^2\right\}$$

$$\leq K\left\{\frac{l_n}{r_n k}i\Delta + \left(\frac{l_n}{r_n}\right)^2 i^2\Delta^2\right\}$$

for K independent of m. Then, applying Móricz' (1982) Theorem in a similar way as for (5.2) in Drees (2000), we get

$$\mathbb{E}\left[\max_{i\in\{0,\ldots,y_0/\Delta\}} |S_m(i\Delta)|^4\right] \leq K\left\{\frac{l_n}{r_n k}\log^4\left(\frac{1}{\Delta}\right) + \left(\frac{l_n}{r_n}\right)^2\right\} \underset{(n\to\infty)}{\longrightarrow} 0,$$

whence (2.43) follows, completing the proof. □

Based upon the result of Theorem 2.4 we can derive a weighted version of the convergence in (2.33):

Theorem 2.5. *Suppose* **(U1)**-**(U3)** *hold and* $q(\cdot)$ *satisfies* (2.13). *Then*

$$\frac{\sqrt{k}}{q(y)}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\{U_i>1-\frac{k}{n}y\}} - \frac{k}{n}y\right]\right\} \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \frac{W(t,y)}{q(y)} \quad \text{in } D^2,$$

where $\{W(t,y)\}$ *is as in Theorem 2.4.*

Proof. For brevity put $e_n(t,y) := \frac{1}{\sqrt{k}}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\{U_i>1-\frac{k}{n}y\}} - \frac{k}{n}y\right]$. In view of Theorem 2.4 and Billingsley (1968, Theorem 4.2), it suffices to prove that for all $\varepsilon > 0$

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} \mathrm{P}\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{e_n(t,y)}{q(y)}\right| > 6\varepsilon\right\} = 0, \tag{2.45}$$

$$\lim_{\vartheta\downarrow 0} \mathrm{P}\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| > \varepsilon\right\} = 0. \tag{2.46}$$

We first show (2.45). For $s = 1,\ldots,n$ define

$$S_s(y) := \frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=1}^{s}\left(I_{\{U_i>1-\frac{k}{n}y\}} - \frac{k}{n}y\right) \qquad \text{and} \qquad \|S_s\| := \sup_{y\in(0,\vartheta]}|S_s(y)|,$$

such that $e_n(t,y)/q(y) = S_{\lfloor nt \rfloor}(y)$. Then

$$
\begin{aligned}
&\mathrm{P}\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{e_n(t,y)}{q(y)}\right| > 3\varepsilon\right\} = \mathrm{P}\left\{\max_{s\in\left\{\lfloor nt_0\rfloor,\ldots,n\right\}}\|S_s\| > 3\varepsilon\right\}\\
&\le \frac{\mathrm{P}\left\{\|S_n\|>\varepsilon\right\} + \mathrm{P}\left\{\max_{\substack{r<s\in\left\{\lfloor nt_0\rfloor,\ldots,n\right\}\\ s-r\le 2r_n}}\|S_s - S_r\| > \varepsilon\right\} + \frac{n}{r_n}\beta(r_n)}{1-\max_{s\in\left\{\lfloor nt_0\rfloor,\ldots,n\right\}}\mathrm{P}\left\{\|S_n - S_s\|>\varepsilon\right\}},
\end{aligned}
\tag{2.47}
$$

where the last step follows from the Ottaviani-type inequality in Bücher (2015, Lemma 3) combined with the fact that α-mixing coefficients are bounded by β-mixing coefficients. Next, we show that the numerator tends to zero and the denominator tends to 1. First consider the three terms in the numerator:

First, because β-mixing coefficients are non-increasing in the argument, we can bound

$$
\frac{n}{r_n}\beta(r_n) \le \frac{n}{r_n}\beta(l_n) \underset{(n\to\infty)}{\overset{\textbf{(U1)}}{=}} o(1).
$$

Second, because of (5.3) in the proof of Drees (2000, Theorem 2.2),

$$
\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty}\mathrm{P}\left\{\|S_n\|>\varepsilon\right\} = \lim_{\vartheta\downarrow 0}\limsup_{n\to\infty}\mathrm{P}\left\{\sup_{y\in(0,\vartheta]}\left|\frac{e_n(1,y)}{q(y)}\right| > \varepsilon\right\} = 0.
$$

Now, for the numerator it remains to be shown that

$$
\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty}\mathrm{P}\left\{\max_{\substack{r<s\in\left\{\lfloor nt_0\rfloor,\ldots,n\right\}\\ s-r\le 2r_n}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right)\right| > \varepsilon\right\} = 0,
$$

where

$$
\begin{aligned}
&\max_{\substack{r<s\in\left\{\lfloor nt_0\rfloor,\ldots,n\right\}\\ s-r\le 2r_n}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right)\right|\\
&\quad\le \underbrace{\max_{\substack{r<s\in\left\{\lfloor nt_0\rfloor,\ldots,n\right\}\\ s-r\le 2r_n}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right|}_{=:A_n} + \underbrace{\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{2r_n}{\sqrt{k}}\frac{k}{n}y\right|}_{=:B_n}.
\end{aligned}
$$

By condition **(U1)** B_n tends to zero. As for A_n

$$
\begin{aligned}
&\mathrm{P}\left\{\max_{\substack{r<s\in\{\lfloor nt_0\rfloor,\dots,n\}\\ s-r\le 2r_n}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}I_{\{U_i>1-\frac{k}{n}y\}}\right|>\varepsilon/2\right\}\\
\le\;&\mathrm{P}\left\{\max_{m\in\{0,\dots,\lfloor n/(2r_n)\rfloor\}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=\lfloor m2r_n\rfloor+1}^{\lfloor(m+2)2r_n\rfloor}I_{\{U_i>1-\frac{k}{n}y\}}\right|>\varepsilon/2\right\}\\
\le\;&\left\lfloor\frac{n}{2r_n}\right\rfloor\mathrm{P}\left\{\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n}I_{\{U_i>1-\frac{k}{n}y\}}\right|>\varepsilon/2\right\}\\
\le\;&\left\lfloor\frac{n}{2r_n}\right\rfloor\sum_{j=0}^{\infty}\mathrm{P}\left\{\sup_{y\in(\vartheta e^{-(j+1)},\vartheta e^{-j}]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n}I_{\{U_i>1-\frac{k}{n}y\}}\right|>\varepsilon/2\right\}\\
\overset{(2.32)}{\le}\;&\left\lfloor\frac{n}{2r_n}\right\rfloor\sum_{j=0}^{\infty}\mathrm{P}\left\{\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n}I_{\{U_i>1-\frac{k}{n}\vartheta e^{-j}\}}>\varepsilon/2q(\vartheta e^{-(j+1)})\right\}\\
\le\;&\left\lfloor\frac{n}{2r_n}\right\rfloor\sum_{j=0}^{\infty}\mathbb{E}\left[\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n}I_{\{U_i>1-\frac{k}{n}\vartheta e^{-j}\}}\right]^2(\varepsilon/2)^{-2}q^{-2}(\vartheta e^{-(j+1)})\\
\le\;&\left\lfloor\frac{n}{2r_n}\right\rfloor\sum_{j=0}^{\infty}K\frac{r_nk}{n}\frac{1}{k}\vartheta e^{-j}(\varepsilon/2)^{-2}q^{-2}(\vartheta e^{-(j+1)})\\
\le\;&K\sum_{j=0}^{\infty}\vartheta e^{-j}q^{-2}(\vartheta e^{-(j+1)}),
\end{aligned}
$$

where the second inequality follows from strict stationarity and the second to last one from Loève's c_r inequality combined with (2.44) and **(U3)**. Using (2.32) the last term can be bounded by

$$
\begin{aligned}
K\sum_{j=0}^{\infty}\left[\vartheta e^{-(j+1)}\right]^{1-2\nu}\left|\log(\vartheta e^{-(j+1)})\right|^{-2\mu}\le& K\int_0^{\infty}\left(\vartheta e^{-t}\right)^{1-2\nu}\left|\log(\vartheta e^{-t})\right|^{-2\mu}\mathrm{d}t\\
=&K\int_{-\log(\vartheta)}^{\infty}e^{-z(1-2\nu)}z^{-2\mu}\mathrm{d}z,
\end{aligned}
$$

which tends to 0 as $\vartheta\downarrow 0$, if and only if $\nu<\frac{1}{2}$ or $\nu=\frac{1}{2}$ and $\mu>\frac{1}{2}$. All in all the numerator tends to zero as $n\to\infty$ followed by $\vartheta\downarrow 0$.

Now consider the denominator of (2.47): by strict stationarity we can write

$$
\begin{aligned}
&\max_{s\in\{\lfloor nt_0\rfloor,\dots,n\}} \mathrm{P}\left\{\|S_n - S_s\| > \varepsilon\right\} \\
&= \max_{s\in\{\lfloor nt_0\rfloor,\dots,n\}} \mathrm{P}\left\{\|S_m\| > \varepsilon\right\} \\
&= \max_{s\in\{\lfloor nt_0\rfloor,\dots,n\}} \mathrm{P}\left\{\sup_{y\in(0,\vartheta]}\left|\frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=1}^{m}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right)\right| > \varepsilon\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
&\frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=1}^{m}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right) \\
&\quad= \underbrace{\frac{1}{q(y)}\sum_{w=0}^{\lfloor m/r_n\rfloor-1}\frac{1}{\sqrt{k}}\sum_{i=wr_n+1}^{(w+1)r_n}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right)}_{=:C_{m,n}} \\
&\qquad\qquad + \underbrace{\frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=\lfloor m/r_n\rfloor r_n+1}^{m}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right)}_{=:D_{m,n}}.
\end{aligned}
$$

We have

$$
D_{m,n} = \underbrace{\frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=\lfloor m/r_n\rfloor r_n+1}^{m} I_{\left\{U_i>1-\frac{k}{n}y\right\}}}_{=:\widetilde{A}_{m,n}} - \underbrace{\frac{m-\lfloor m/r_n\rfloor r_n}{\sqrt{k}}\frac{k}{n}\frac{y}{q(y)}}_{=:\widetilde{B}_{m,n}}.
$$

Because there are at most r_n terms in the sum in $\widetilde{A}_{m,n}$, that $\sup_{y\in(0,\vartheta]}\widetilde{A}_{m,n} = o_{\mathrm{P}}(1)$ uniformly in m can be seen as for A_n. The convergence of $\sup_{y\in(0,\vartheta]}\widetilde{B}_{m,n}$ (uniformly in m) can also be seen as the one for B_n. It remains to investigate $C_{m,n}$. To this end consider the proof of Theorem 2.2 in Drees (2000). Replacing his $m_n = \left\lfloor\frac{n}{2r_n}\right\rfloor$ by $m_n = \left\lfloor\frac{m}{2r_n}\right\rfloor$ in the proof it is easy to see that

$$
\lim_{\vartheta\downarrow 0}\lim_{n\to\infty}\mathrm{P}\left\{\sup_{y\in\left(0,\vartheta e^{-j_n}\right]}\left|C_{m,n}\right| > \varepsilon\right\} = 0 \quad \text{uniformly in } m\in\{1,\dots,n\},
$$

where $j_n := \min\{j \in \mathbb{N} : \sqrt{k} \le \eta \frac{q}{Id}(\vartheta e^{-(j+1)})\}$ for some small $\eta > 0$. The uniformity is due to the fact that for all $m \in \{1, \dots, n\}$ in Drees' (2000) notation

$$\mathbb{E}\left(\widetilde{S}_n(\vartheta e^{-j})\right) \le \sqrt{k}\vartheta e^{-j}$$

in the step leading to his (5.6). Using assumption **(U3)** and again replacing $m_n = \left\lfloor \frac{n}{2r_n} \right\rfloor$ by $m_n = \left\lfloor \frac{m}{2r_n} \right\rfloor$ in the proof of Drees (2000, Theorem 2.3), retracing the proof again yields

$$\lim_{\vartheta \downarrow 0} \lim_{n\to\infty} \mathrm{P}\left\{ \sup_{y\in\left(\vartheta e^{-j_n},\vartheta\right]} |C_{m,n}| > \varepsilon \right\} = 0 \quad \text{uniformly in } m \in \{1, \dots, n\}\,.$$

The uniformity is due to the uniformity of his moment inequality (5.14) derived there by an application of Burkholder's inequality.

Next we will prove (2.46) via Lin and Choi (1999, Lemma 2.1). Use that **(U3)** implies

$$\begin{aligned}\frac{n}{r_n k} \operatorname{Var}\left(\sum_{i=1}^{r_n} I_{\left\{1-\frac{k}{n}y_2<U_i\le 1-\frac{k}{n}y_1\right\}}\right) &\le \frac{n}{r_n k}\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{1-\frac{k}{n}y_2<U_i\le 1-\frac{k}{n}y_1\right\}}\right]^4 \\ &\le C(y_2 - y_1).\end{aligned}$$

(Recall again for the first inequality that the sum of the indicators is $\mathbb{N}_0$-valued.) By **(U2)** the left-hand side converges to $r(y_2, y_2) - 2r(y_1, y_2) + r(y_1, y_1)$ as $n \to \infty$. Hence,

$$r(y_2, y_2) - 2r(y_1, y_2) + r(y_1, y_1) \le C\,|y_1 - y_2|\,. \tag{2.48}$$

Assume w.l.o.g. $t_1 > t_2$ and use the Cauchy-Schwarz inequality in the last step to obtain

$$\begin{aligned} & \mathbb{E}\left[W(t_1, y_1) - W(t_2, y_2)\right]^2 \\ = \;& \operatorname{Var}(W(t_1, y_1)) + \operatorname{Var}(W(t_2, y_2)) - 2\operatorname{Cov}(W(t_1, y_1), W(t_2, y_2)) \\ = \;& t_1 r(y_1, y_1) + t_2 r(y_2, y_2) - 2\min(t_1, t_2) r(y_1, y_2) \\ = \;& (t_1 - t_2) r(y_1, y_1) + t_2\left\{r(y_2, y_2) - 2r(y_1, y_2) + r(y_1, y_1)\right\} \\ \le \;& C\left\{|t_1 - t_2| + |y_1 - y_2|\right\} \\ \le \;& \sqrt{2}C\left\{|t_1 - t_2|^2 + |y_1 - y_2|^2\right\}^{1/2} =: \varphi^2\left(\left\|\begin{pmatrix}t_1\\ y_1\end{pmatrix} - \begin{pmatrix}t_2\\ y_2\end{pmatrix}\right\|_2\right), \end{aligned}$$

i.e., $\varphi(r) = \sqrt{2}C\sqrt{r}$. Next, define

$$\begin{aligned}
&\mathbb{D}_j := [t_0, 1] \times \left[\vartheta e^{-(j+1)}, \vartheta e^{-j}\right];\\
&\Gamma_j^2 := \sup_{(t,y)\in\mathbb{D}_j} \mathbb{E}\left[W(t,y)\right]^2 = \sup_{y\in\left[\vartheta e^{-(j+1)},\vartheta e^{-j}\right]} r(y,y) \overset{(2.48)}{\leq} C\vartheta e^{-(j+1)},\\
&\lambda_j := \vartheta e^{-(j+1)}\left[e-1\right],
\end{aligned}$$

so that

$$\int_0^\infty \varphi(\sqrt{2}\lambda_j 2^{-x^2})\mathrm{d}x \leq K\sqrt{\lambda_j}\int_0^\infty 2^{-x^2/2}\mathrm{d}x = \mathcal{O}\left(\vartheta^{1/2}e^{-\frac{1}{2}(j+1)}\right),$$

and apply Lemma 2.1 of Lin and Choi (1999) to get, using $\nu < \frac{1}{2}$ or $\nu = \frac{1}{2}$ and $\mu > 1/2$ from (2.13),

$$\begin{aligned}
&\mathrm{P}\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| > \varepsilon\right\}\\
\leq\ & \sum_{j=0}^\infty \mathrm{P}\left\{\sup_{\substack{t\in[t_0,1]\\ y\in\left(\vartheta e^{-(j+1)},\vartheta e^{-j}\right]}} \left|W(t,y)\right| > \varepsilon\left(\vartheta e^{-(j+1)}\right)^\nu\left|\log(\vartheta e^{-(j+1)})\right|^\mu\right\}\\
\leq\ & K\sum_{j=0}^\infty \exp\left\{-\frac{1}{2}\left(\frac{\varepsilon\left(\vartheta e^{-(j+1)}\right)^\nu\left|\log(\vartheta e^{-(j+1)})\right|^\mu}{\Gamma_j + (2\sqrt{2}+2)K_1\int_0^\infty\varphi(\sqrt{2}\lambda_j 2^{-x^2})\mathrm{d}x}\right)^2\right\}\\
\leq\ & K\sum_{j=0}^\infty \exp\left\{-K_1\left(\vartheta e^{-(j+1)}\right)^{2\nu-1}\left|\log(\vartheta e^{-(j+1)})\right|^{2\mu}\right\}\\
\leq\ & K\vartheta^{K_2}\sum_{j=0}^\infty \exp\left\{-K_2\right\}^{j+1} \underset{(\vartheta\downarrow 0)}{\longrightarrow} 0,
\end{aligned} \tag{2.49}$$

by boundedness of the sum. □

Proposition 2.2. *Suppose $q(\cdot)$ and $W(\cdot,\cdot)$ are as in Theorem 2.5. Then*

$$\lim_{\vartheta\downarrow 0}\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| \overset{\text{a.s.}}{=} 0.$$

Proof. The proof of (2.49) reveals that by choosing $\vartheta = \vartheta_n = n^{-2/K_2}$ one obtains

$$\sum_{n=1}^{\infty} \mathrm{P}\left\{ \sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta_n]}} \left| \frac{W(t,y)}{q(y)} \right| > \varepsilon \right\} < \infty,$$

which implies by the Borel-Cantelli lemma that, a.s.,

$$\lim_{\vartheta\downarrow 0} \sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}} \left| \frac{W(t,y)}{q(y)} \right| = \limsup_{\vartheta\downarrow 0} \sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}} \left| \frac{W(t,y)}{q(y)} \right| = 0,$$

because $\varepsilon > 0$ was arbitrary. □

Proof of Theorem 2.1: By the proof of Theorem 3.1 in Drees (2000) **(A1)-(A4)** imply **(U1)-(U3)** for $U_i := F(X_i) \sim \mathcal{U}[0,1]$. Hence, noting that

$$X_i > U\left(\frac{n}{ky}\right) \quad \Longleftrightarrow \quad U_i > 1 - \frac{k}{n}y,$$

we obtain from Theorem 2.5 that

$$\frac{\sqrt{k}}{q(y)} \left(\frac{1}{k} \sum_{i=1}^{\lfloor nt \rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - ty \right) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{1}{q(y)} W(t,y) \quad \text{in } D^2. \tag{2.50}$$

Applying the continuous mapping theorem to (2.50), we get

$$\frac{\sqrt{k}}{q(y)} \begin{pmatrix} \frac{1}{k}\sum_{i=1}^{\lfloor nt \rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - ty \\ \frac{1}{k}\sum_{i=\lfloor nt \rfloor+1}^{n} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - (1-t)y \\ \frac{1}{k}\sum_{i=\lfloor nt \rfloor+1}^{\lfloor n(t+t_0) \rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - t_0 y \end{pmatrix} \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{1}{q(y)} \begin{pmatrix} W(t,y) \\ W(1,y) - W(t,y) \\ W(t+t_0,y) - W(t,y) \end{pmatrix}$$

in $D^3([t_0, 1-t_0] \times [0, y_0+\delta])$. By Skorohod's representation theorem (cf., e.g., Wichura, 1970, Theorem 1) we can pretend that this convergence holds almost surely on a suitable probability space:

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,y_0+\delta]}} \frac{1}{q(y)}\left|\sqrt{k}\begin{pmatrix} \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>U\left(\frac{n}{ky}\right)\right\}} - ty \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{n} I_{\left\{X_i>U\left(\frac{n}{ky}\right)\right\}} - (1-t)y \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{\lfloor n(t+t_0)\rfloor} I_{\left\{X_i>U\left(\frac{n}{ky}\right)\right\}} - t_0 y \end{pmatrix} - \begin{pmatrix} W(t,y) \\ W(1,y)-W(t,y) \\ W(t+t_0,y)-W(t,y) \end{pmatrix}\right| \underset{(n\to\infty)}{\xrightarrow{\text{a.s.}}} 0. \quad (2.51)$$

(Note that the limits are continuous, hence convergence is uniform.) It remains to show that $U\left(\frac{n}{ky}\right)$ can be replaced by $y^{-\gamma}U\left(\frac{n}{k}\right)$ in (2.51). For brevity we carry out the steps for the first component of (2.51) only (the others being dealt with similarly). Similarly as in the proof of Corollary 3 in Einmahl *et al.* (2016) we set

$$y_n := \frac{n}{k}\left[1 - F\left(y^{-\gamma}U\left(\frac{n}{k}\right)\right)\right], \quad y \in (0, y_0+\delta],$$

so that by **(A4)** (cf. Einmahl *et al.*, 2016, p. 46)

$$\sup_{y\in(0,y_0+\delta]}\left|\frac{y_n - y}{A(n/k)y}\right| \underset{(n\to\infty)}{=} \mathcal{O}(1). \quad (2.52)$$

Inserting y_n for y in the first component of (2.51) gives

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,y_0+\delta]}} \frac{1}{q(y_n)}\left|\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>y^{-\gamma}U\left(\frac{n}{k}\right)\right\}} - ty_n\right) - W(t,y_n)\right| \underset{(n\to\infty)}{\xrightarrow{\text{a.s.}}} 0 \quad (2.53)$$

Now we have to show that y_n can be replaced with y at the three occurences in (2.53). For $q(\cdot)$, by the first property in (2.13), it suffices to note that (using $y_n \underset{(n\to\infty)}{=} y(1+o(1))$ uniformly in y from (2.52))

$$\begin{aligned}\sup_{y\in\left(0,\vartheta/2\right]}\left|\frac{q(y_n)}{q(y)}\right| &\overset{(2.32)}{=} \sup_{y\in\left(0,\vartheta/2\right]} \frac{y_n^{\nu}\,|\log y_n|^{\mu}}{y^{\nu}\,|\log y|^{\mu}} \\ &= (1+o(1))^{\nu}\sup_{y\in\left(0,\vartheta/2\right]}\left|\frac{\log(y)+\log(1+o(1))}{\log(y)}\right|^{\mu} \underset{(n\to\infty)}{=} 1+o(1). \quad (2.54)\end{aligned}$$

Combine (2.52) with $\sqrt{k}A(n/k) \to 0$ to see that ty_n may be replaced with ty. Finally, by a simple (uniform) continuity argument we have that for all $\vartheta > 0$

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in[\vartheta,y_0+\delta]}} \frac{1}{q(y)} \left|W(t,y_n) - W(t,y)\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \tag{2.55}$$

Further, by Proposition 2.2, (2.52) and (2.54)

$$\begin{aligned}
&\sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,\vartheta]}} \frac{1}{q(y)} \left|W(t,y_n) - W(t,y)\right| \\
&\qquad \le \sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,\vartheta]}} \frac{q(y_n)}{q(y)} \left|\frac{W(t,y_n)}{q(y_n)}\right| + \sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,\vartheta]}} \left|\frac{W(t,y)}{q(y)}\right| \overset{\text{a.s.}}{\longrightarrow} 0,
\end{aligned}$$

as $\vartheta \downarrow 0$ and $n \to \infty$, justifying the replacement in $W(\cdot,\cdot)$. □

Proof of Corollary 2.1: We will only prove the convergence of the first component in (2.15), the others being proved similarly. Theorem 2.1 implies (because $y_0 \ge 1$)

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in[1/2,1+\delta]}} \left|\sqrt{k}\left\{\frac{1}{kt}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > y^{-\gamma}U(n/k)\}} - y\right\} - \frac{W(t,y)}{t}\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0.$$

It follows exactly as in the proof of Einmahl *et al.* (2016, Theorem 3) using a generalized Vervaat lemma (cf. Einmahl, Gantner and Sawitzki, 2010, Lemma 5) that

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in[1/2,1]}} \left|\sqrt{k}\left\{\left(\frac{X_k(0,t,y)}{U(n/k)}\right)^{-1/\gamma} - y\right\} + \frac{W(t,y)}{t}\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0, \tag{2.56}$$

so in particular

$$\sup_{t\in[t_0,1-t_0]} \left|\sqrt{k}\left\{\left(\frac{X_k(0,t,1)}{U(n/k)}\right)^{-1/\gamma} - 1\right\} + \frac{W(t,1)}{t}\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \tag{2.57}$$

Replacing y with $y_n := yX_k(0,t,1)/U(n/k)$ in Theorem 2.1 implies, using (2.57),

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}}} \frac{1}{q\left(y_n^{-1/\gamma}\right)} \left| \sqrt{k}\left\{ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>yX_k(0,t,1)\right\}} - y_n^{-1/\gamma}t \right\} - W\left(t, y_n^{-1/\gamma}\right)\right| \underset{(n\to\infty)}{\xrightarrow{\text{a.s.}}} 0. \tag{2.58}$$

Now we show that $y_n^{-1/\gamma}$ can be replaced by $y^{-1/\gamma}$ at the three occurrences in (2.58). By the first property in (2.13), we need only justify the replacement in $q(\cdot)$ for large y. Since by (2.57) $y_n^{-1/\gamma} \underset{(n\to\infty)}{\overset{\text{a.s.}}{=}} y^{-1/\gamma}(1+o(1))$ uniformly in t, y, it follows as in (2.54) that

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq(\vartheta/2)^{-\gamma}}} \left| \frac{q(y_n^{-1/\gamma})}{q\left(y^{-1/\gamma}\right)} \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{=}} 1+o(1).$$

As for $W(\cdot,\cdot)$, the same arguments as in (2.55) and below apply. Last, by (2.57) uniformly in t (and y)

$$\frac{1}{q\left(y^{-1/\gamma}\right)} t\sqrt{k}\left(y_n^{-1/\gamma} - y^{-1/\gamma}\right) \underset{(n\to\infty)}{\overset{\text{a.s.}}{=}} -\frac{1}{q\left(y^{-1/\gamma}\right)} y^{-1/\gamma}W(t,1) + o(1).$$

Making the replacements in (2.58) and multiplying through with $k/\lfloor kt\rfloor$ yields

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}}} \frac{1}{q\left(y^{-1/\gamma}\right)} \left| \sqrt{k}\left(F_n(0,t,y) - \frac{kt}{\lfloor kt\rfloor} y^{-1/\gamma}\right) - \right.$$
$$\left. \frac{k}{\lfloor kt\rfloor}\left(W\left(t, y^{-1/\gamma}\right) - y^{-1/\gamma}W(t,1)\right)\right| \underset{(n\to\infty)}{\xrightarrow{\text{a.s.}}} 0.$$

Using $\frac{k}{\lfloor kt\rfloor} = 1/t + \mathcal{O}(1/k)$ uniformly in t and Proposition 2.2, the conclusion follows. □

Proof of Proposition 2.1: We will derive convergence of $t\sqrt{k}(\widehat{\gamma}_{WLS}(0,t)-\gamma)$ from the first component of (2.15), the convergences of $(1-t)\sqrt{k}(\widehat{\gamma}_{WLS}(t,1)-\gamma)$ and $t_0\sqrt{k}(\widehat{\gamma}_{WLS}(t,t+t_0)-\gamma)$ following similarly from the other components. The required joint convergence then follows from the joint convergence in (2.15).

Noting that for $i = 0, 1, \ldots, \lfloor kt\rfloor - 1$

$$F_n(0,t,y) = \frac{\lfloor kt\rfloor - i}{\lfloor kt\rfloor} \quad \text{constant on } y \in \left[\frac{X_{\lfloor nt\rfloor-\lfloor kt\rfloor+i:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor}}, \frac{X_{\lfloor nt\rfloor-\lfloor kt\rfloor+i+1:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor}}\right),$$

$$F_n(0,t,y) = 0 \qquad\qquad \text{for } y \geq \frac{X_{\lfloor nt \rfloor : \lfloor nt \rfloor}}{X_{\lfloor nt \rfloor - \lfloor kt \rfloor : \lfloor nt \rfloor}},$$

it is straightforward to check with **(W1)** that

$$\begin{aligned}\int_1^\infty \left\{ \int_0^{F_n(0,t,y)} J(s)\mathrm{d}s \right\} \frac{\mathrm{d}y}{y} &= \sum_{i=0}^{\lfloor kt \rfloor - 1} \left\{ \int_0^{\frac{\lfloor kt \rfloor - i}{\lfloor kt \rfloor}} J(s)\mathrm{d}s \right\} \log \left(\frac{X_{\lfloor nt \rfloor - \lfloor kt \rfloor + i + 1 : \lfloor nt \rfloor}}{X_{\lfloor nt \rfloor - \lfloor kt \rfloor + i : \lfloor nt \rfloor}} \right) \\ &= \widehat{\gamma}_{WLS}(0,t).\end{aligned}$$

Using **(W2)**, **(W3)** and partial integration it is further easy to establish that

$$\int_1^\infty \left\{ \int_0^{y^{-1/\gamma}} J(s)\mathrm{d}s \right\} \frac{\mathrm{d}y}{y} = \gamma \int_0^1 \left\{ \int_0^z J(s)\mathrm{d}s \right\} \frac{\mathrm{d}z}{z} = \gamma.$$

Combine these two facts to obtain

$$\begin{aligned}\sqrt{k}(\widehat{\gamma}_{WLS}(0,t) - \gamma) &= \sqrt{k} \int_1^\infty \left\{ \int_0^{F_n(0,t,y)} J(s)\mathrm{d}s - \int_0^{y^{-1/\gamma}} J(s)\mathrm{d}s \right\} \frac{\mathrm{d}y}{y} \\ &= \gamma\sqrt{k} \int_0^1 \left\{ \int_0^{F_n(0,t,u^{-\gamma})} J(s)\mathrm{d}s - \int_0^u J(s)\mathrm{d}s \right\} \frac{\mathrm{d}u}{u}. \qquad (2.59)\end{aligned}$$

Write the result of Corollary 2.1 as

$$\sup_{\substack{t \in [t_0, 1-t_0] \\ u \in (0,1]}} \frac{1}{q(u)} \left| \sqrt{k} \left(F_n(0,t,u^{-\gamma}) - u \right) - \frac{1}{t} \left[W(t,u) - uW(t,1) \right] \right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0. \qquad (2.60)$$

Using the mean value theorem for the function $x \mapsto \int_0^x J(s)\mathrm{d}s$, we get for some $\xi = \xi_u \in (0,1)$

$$\int_0^{F_n(0,t,u^{-\gamma})} J(s)\mathrm{d}s = \int_0^u J(s)\mathrm{d}s + \left(F_n(0,t,u^{-\gamma}) - u \right) J\left(u + \xi \left(F_n(0,t,u^{-\gamma}) - u \right) \right).$$

Thus, from (2.59) and (2.60), uniformly in $t \in [t_0, 1-t_0]$

$$\begin{aligned}&\sqrt{k}(\widehat{\gamma}_{WLS}(0,t) - \gamma) \\ &= \gamma\sqrt{k} \int_0^1 \left(F_n(0,t,u^{-\gamma}) - u \right) J\left(u + \xi \left(F_n(0,t,u^{-\gamma}) - u \right) \right) \frac{\mathrm{d}u}{u}\end{aligned}$$

$$\overset{\text{a.s.}}{=} \frac{\gamma}{t} \int_0^1 \left(W(t,u) - uW(t,1) + o(1)q(u)\right) J\left(u + \xi\left(F_n(0,t,u^{-\gamma}) - u\right)\right) \frac{\mathrm{d}u}{u} \tag{2.61}$$

$$\underset{(n\to\infty)}{\longrightarrow} \frac{\gamma}{t} \int_0^1 \left(W(t,u) - uW(t,1)\right) J(u) \frac{\mathrm{d}u}{u} \overset{\textbf{(W1)}}{=} \frac{\gamma}{t} \int_0^1 W(t,u) J(u) \frac{\mathrm{d}u}{u}.$$

Note that the integral in (2.61) is well-defined because of Proposition 2.2.

By calculating covariances of $\gamma \int_0^1 W(t,u)J(u) \frac{\mathrm{d}u}{u}$ we conclude (going back to the original probability space)

$$t\sqrt{k}(\widehat{\gamma}_{WLS}(0,t) - \gamma) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \sigma_{\widehat{\gamma}_{WLS},\gamma} W(t) \quad \text{in } D[t_0, 1-t_0],$$

where $W(\cdot)$ is a standard Brownian motion and

$$\sigma^2_{\widehat{\gamma}_{WLS},\gamma} = \gamma^2 \int_0^1 \int_0^1 \frac{r(x,y)}{xy} J(x)J(y)\mathrm{d}x\mathrm{d}y.$$

□

Proof of Theorem 2.2: We only prove part (a), the proof of (b) being similar. Because t_0 can be chosen arbitrarily close to 0 in (2.22), similarly as in the proof of Drees (2003, Theorem 2.3), one obtains (on a suitable probability space) via a diagonal argument that

$$\sup_{t\in[t_n,1]} \left| \sqrt{k}\left(\widehat{\gamma}(0,t) - \gamma\right) - \sigma_{\widehat{\gamma},\gamma} \frac{W(t)}{t} \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0,$$

for some sequence $t_n \downarrow 0$ tending to zero not too fast, whence with $\widetilde{t}_n := \frac{\lfloor nt_n \rfloor + 1}{n} (\geq t_n)$

$$\begin{aligned}
\frac{k}{n} \sum_{i=\lfloor nt_n \rfloor + 1}^{n} \left[\widehat{\gamma}(0,i/n) - \widehat{\gamma}(0,1)\right]^2 &= \int_{\widetilde{t}_n}^1 k\left(\widehat{\gamma}(0,t) - \widehat{\gamma}(0,1)\right)^2 \mathrm{d}t \\
&= \int_{\widetilde{t}_n}^1 \left[\sqrt{k}\left(\widehat{\gamma}(0,t) - \gamma\right) - \sqrt{k}\left(\widehat{\gamma}(0,1) - \gamma\right)\right]^2 \mathrm{d}t \\
&\overset{\text{a.s.}}{=} \sigma^2_{\widehat{\gamma},\gamma} \int_{\widetilde{t}_n}^1 \left(\frac{W(t)}{t} - W(1)\right)^2 \mathrm{d}t(1+o(1)) \\
&= \sigma^2_{\widehat{\gamma},\gamma} \int_{\log(\widetilde{t}_n)}^0 \left(\frac{W(e^y)}{e^{y/2}} - e^{y/2}W(1)\right)^2 \mathrm{d}y(1+o(1)),
\end{aligned} \tag{2.62}$$

using the substitution $t = e^y$ ($\mathrm{d}t = e^y \mathrm{d}y$) in the fourth equality.

Now observe that $W(e^\cdot)/e^{\cdot/2}$ is a zero-mean Gaussian process with covariance function

$$\mathrm{E}\left[\frac{W(e^x)}{e^{x/2}}\frac{W(e^y)}{e^{y/2}}\right] = e^{-(x+y)/2}\min(e^x, e^y) = e^{\min(x-y,y-x)/2} = e^{-|x-y|/2}$$

only depending on $x - y$, which implies (cf. Cramér and Leadbetter, 1967, p. 122) strict stationarity. By an application of the Birkhoff-Khintchine ergodic theorem (cf. Cramér and Leadbetter, 1967, p. 151) we obtain

$$\frac{1}{\log(1/\widetilde{t}_n)}\sigma^2_{\widehat{\gamma},\gamma}\int_{\log(\widetilde{t}_n)}^0 \left(\frac{W(e^y)}{e^{y/2}}\right)^2 \mathrm{d}y \xrightarrow[(n\to\infty)]{\text{a.s.}} \sigma^2_{\widehat{\gamma},\gamma}\,\mathrm{E}\left(\frac{W(e^y)}{e^{y/2}}\right)^2 = \sigma^2_{\widehat{\gamma},\gamma}. \tag{2.63}$$

Noting that $\int_{\log(\widetilde{t}_n)}^0 \left(e^{y/2}W(1)\right)^2 \mathrm{d}y = \mathcal{O}(1)$, the conclusion follows from (2.62) and (2.63). $\square$

Proof of Theorem 2.3: We will focus on the result under $\mathcal{H}_1^>$ (i.e., $\gamma_{\text{pre}} > \gamma_{\text{post}}$, meaning heavier pre-break tails) as the other can be established similarly. Write $X_i = X_{i,n}$ and $y_0 = y_{0,\text{pre}}$ for brevity. Then Theorem 2.1 implies for some $\widetilde{\delta} > 0$ that may change from line to line in this proof

$$\frac{1}{q(y)}\sqrt{k_{\text{pre}}}\left(\frac{1}{k_{\text{pre}}}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i > y^{-\gamma_{\text{pre}}}U_{\text{pre}}(n/k_{\text{pre}})\right\}} - yt\right) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{W_{\text{pre}}(t,y)}{q(y)}, \tag{2.64}$$

in $D([t_0, t^*] \times [0, y_0 + \widetilde{\delta}])$ and for the post-break r.v.s

$$\begin{aligned}\sup_{y\geq y_0^{-\gamma_{\text{pre}}}-\widetilde{\delta}} \Bigg| \sqrt{k_{\text{post}}}\left(\frac{1}{k_{\text{post}}}\sum_{i=\lfloor nt^*\rfloor+1}^{n} I_{\left\{X_i > yU_{\text{post}}(n/k_{\text{post}})\right\}} - y^{-1/\gamma_{\text{post}}}(1-t^*)\right) \\ - \left(W_{\text{post}}\left(1, y^{-1/\gamma_{\text{post}}}\right) - W_{\text{post}}\left(t^*, y^{-1/\gamma_{\text{post}}}\right)\right)\Bigg| \Big/ q\left(y^{-1/\gamma_{\text{post}}}\right) \xrightarrow[(n\to\infty)]{\text{a.s.}} 0.\end{aligned} \tag{2.65}$$

Inserting $y_n := yU_{\text{pre}}(n/k_{\text{pre}})/U_{\text{post}}(n/k_{\text{post}})$ for y in (2.65) and recalling Proposition 2.2 gives

$$\sup_{y\geq y_0^{-\gamma_{\text{pre}}}-\widetilde{\delta}} \Bigg| \sqrt{k_{\text{post}}}\Bigg(\frac{1}{k_{\text{post}}}\sum_{i=\lfloor nt^*\rfloor+1}^{n} I_{\left\{X_i > yU_{\text{pre}}(n/k_{\text{pre}})\right\}} -$$

$$y_n^{-1/\gamma_{\text{post}}}(1-t^*)\Bigg)\Bigg| \Big/ q\left(y_n^{-1/\gamma_{\text{post}}}\right) \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \quad (2.66)$$

Further, $y_n^{-1/\gamma_{\text{post}}}(1-t^*)$ may be omitted in (2.66), since, by (2.26) and (2.32), for n large

$$\sqrt{k_{\text{post}}} \sup_{y \geq y_0^{-\gamma_{\text{pre}}}-\widetilde{\delta}} \left| \frac{y_n^{-1/\gamma_{\text{post}}}}{q\left(y_n^{-1/\gamma_{\text{post}}}\right)} \right| \leq \sqrt{k_{\text{post}}} y_n^{-1/(2\gamma_{\text{post}})} \underset{(n\to\infty)}{=} o(1).$$

Using $k_{\text{post}} = \mathcal{O}\left(k_{\text{pre}}\right)$ and, by (2.13) and (2.26), for n sufficiently large

$$q\left(y_n^{-1/\gamma_{\text{post}}}\right) / q\left(y^{-1/\gamma_{\text{pre}}}\right) \leq 1,$$

this yields

$$\sup_{y \geq y_0^{-\gamma_{\text{pre}}}-\widetilde{\delta}} \frac{1}{q\left(y^{-1/\gamma_{\text{pre}}}\right)} \left| \sqrt{k_{\text{pre}}} \left(\frac{1}{k_{\text{pre}}} \sum_{i=\lfloor nt^* \rfloor+1}^{n} I_{\left\{X_i > y U_{\text{pre}}(n/k_{\text{pre}})\right\}} \right) \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0.$$

Going back to the original probability space we obtain by non-negativity of the indicators

$$\frac{\sqrt{k_{\text{pre}}}}{q\left(y\right)} \left(\frac{1}{k_{\text{pre}}} \sum_{i=\lfloor nt^* \rfloor+1}^{\lfloor nt \rfloor} I_{\left\{X_i > y^{-\gamma_{\text{pre}}} U_{\text{pre}}(n/k_{\text{pre}})\right\}} \right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} 0 \quad \text{in } D([t^*, 1-t_0] \times [0, y_0+\widetilde{\delta}]). \tag{2.67}$$

Hence, letting $k = k_{\text{pre}}$ for brevity in the rest of the proof, we get from (2.64) and (2.67) that

$$\frac{\sqrt{k}}{q\left(y\right)} \left(\frac{1}{k} \sum_{i=1}^{\lfloor nt \rfloor} I_{\left\{X_i > y^{-\gamma_{\text{pre}}} U_{\text{pre}}(n/k)\right\}} - y^{-1/\gamma_{\text{pre}}} \min(t^*, t)) \right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \frac{W_{\text{pre}}\left(\min(t^*, t), y\right)}{q\left(y\right)}$$

in $D([t_0, 1-t_0] \times [0, y_0+\widetilde{\delta}])$ or, invoking a Skorohod construction again and putting $t_{\min} := \min(t, t^*)$,

$$\sup_{\substack{t \in [t_0, 1-t_0] \\ y \in (0, y_0+\widetilde{\delta}]}} \frac{1}{q\left(y\right)} \left| \sqrt{k} \left(\frac{1}{k t_{\min}} \sum_{i=1}^{\lfloor nt \rfloor} I_{\left\{X_i > y^{-\gamma_{\text{pre}}} U_{\text{pre}}(n/k)\right\}} - y \right) - \frac{W_{\text{pre}}\left(t_{\min}, y\right)}{t_{\min}} \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \tag{2.68}$$

Then, similarly as for (2.56), it follows that

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in[1/2,y_0]}}\left|\sqrt{k}\left(\left(\frac{X_k\left(0,t,\frac{t_{\min}}{t}y\right)}{U_{\text{pre}}(n/k)}\right)^{-1/\gamma_{\text{pre}}}-y\right)+\frac{W_{\text{pre}}\left(t_{\min},y\right)}{t_{\min}}\right|\xrightarrow[(n\to\infty)]{\text{a.s.}}0,$$

which, for $y=t/t_{\min}\le(1-t_0)/t_0=y_0$, implies

$$\sup_{t\in[t_0,1-t_0]}\left|\sqrt{k}\left(\left(\frac{X_k\left(0,t,1\right)}{U_{\text{pre}}(n/k)}\right)^{-1/\gamma_{\text{pre}}}-\frac{t}{t_{\min}}\right)+\frac{W_{\text{pre}}\left(t_{\min},\frac{t}{t_{\min}}\right)}{t_{\min}}\right|\xrightarrow[(n\to\infty)]{\text{a.s.}}0. \quad (2.69)$$

Substituting $\left(\frac{X_k(0,t,1)}{U_{\text{pre}}(n/k)}y\right)^{-1/\gamma_{\text{pre}}}$ $(y\in[1-\widetilde{\delta},\infty))$ for y in (2.68) thus yields

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\ge1-\widetilde{\delta}}}\frac{1}{q\left(\left(\frac{X_k(0,t,1)}{U_{\text{pre}}(n/k)}y\right)^{-1/\gamma_{\text{pre}}}\right)}\left|\sqrt{k}\left(\frac{1}{kt_{\min}}\sum_{i=1}^{\lfloor nt\rfloor}I_{\left\{X_i>yX_k(0,t,1)\right\}}-y^{-1/\gamma_{\text{pre}}}\left(\frac{X_k\left(0,t,1\right)}{U_{\text{pre}}(n/k)}\right)^{-1/\gamma_{\text{pre}}}\right)\right.$$
$$\left.-\frac{1}{t_{\min}}W_{\text{pre}}\left(t_{\min},\left(\frac{X_k\left(0,t,1\right)}{U_{\text{pre}}(n/k)}y\right)^{-1/\gamma_{\text{pre}}}\right)\right|\xrightarrow[(n\to\infty)]{\text{a.s.}}0. \quad (2.70)$$

As in the proof of Corollary 2.1 it follows from (2.69) and (2.70) that

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\ge1}}\frac{1}{q\left(\left(y\frac{t}{t_{\min}}\right)^{-1/\gamma_{\text{pre}}}\right)}\left|\sqrt{k}\left(\frac{1}{kt_{\min}}\sum_{i=1}^{\lfloor nt\rfloor}I_{\left\{X_i>yX_k(0,t,1)\right\}}-y^{-1/\gamma_{\text{pre}}}\frac{t}{t_{\min}}\right)\right.$$
$$\left.-\frac{1}{t_{\min}}\left[W_{\text{pre}}\left(t_{\min},y^{-1/\gamma_{\text{pre}}}\frac{t}{t_{\min}}\right)-y^{-1/\gamma_{\text{pre}}}W_{\text{pre}}\left(t_{\min},\frac{t}{t_{\min}}\right)\right]\right|\xrightarrow[(n\to\infty)]{\text{a.s.}}0,$$

or, using

$$0\le\frac{q\left(y^{-1/\gamma_{\text{pre}}}\right)}{q\left(\left(y\frac{t}{t_{\min}}\right)^{-1/\gamma_{\text{pre}}}\right)}\le\frac{q\left(y^{-1/\gamma_{\text{pre}}}\right)}{q\left(\left(y\frac{1-t_0}{t_0}\right)^{-1/\gamma_{\text{pre}}}\right)}\le K$$

for y sufficiently large, because $y\mapsto q(y^{-1/\gamma_{\text{pre}}})$ is $RV_{-\nu/\gamma_{\text{pre}}}$ (recall (2.32)),

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq 1}} \frac{1}{q\left(y^{-1/\gamma_{\mathrm{pre}}}\right)}\left|\sqrt{k}\left(\frac{1}{kt}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>yX_k(0,t,1)\right\}} - y^{-1/\gamma_{\mathrm{pre}}}\right)\right.$$
$$\left. -\frac{1}{t}\left[W_{\mathrm{pre}}\left(t_{\min}, y^{-1/\gamma_{\mathrm{pre}}}\frac{t}{t_{\min}}\right) - y^{-1/\gamma_{\mathrm{pre}}}W_{\mathrm{pre}}\left(t_{\min}, \frac{t}{t_{\min}}\right)\right]\right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0.$$

Using this result the convergence in (2.28) can be checked easily by following the derivations in Examples 2.3-2.5. □

Proof of Corollary 2.3: For (a) ((b) being proved similarly) Theorem 2.3 implies the convergence in (2.28), where by calculating covariances the following distributional equality holds

$$\left\{B_{\mathrm{pre}}(t)\right\}_{t\in[0,t^*]} \overset{\mathcal{D}}{=} \left\{\sigma_{\widehat{\gamma},\gamma_{\mathrm{pre}}} W(t)\right\}_{t\in[0,t^*]}, \tag{2.71}$$

where $\sigma^2_{\widehat{\gamma},\gamma_{\mathrm{pre}}}$ is either (2.17), (2.19) or (2.21) with γ, $r(\cdot,\cdot)$ replaced by γ_{pre}, $r_{\mathrm{pre}}(\cdot,\cdot)$ and $W(\cdot)$ is a standard Brownian motion. Write, using $\widetilde{t}_n = \frac{\lfloor nt_n\rfloor+1}{n}$ as defined in the proof of Theorem 2.2,

$$\begin{aligned}
&\int_{\widetilde{t}_n}^{1} k\left(\widehat{\gamma}(0,t)-\widehat{\gamma}(0,1)\right)^2 \mathrm{d}t \\
&= \int_{\widetilde{t}_n}^{t^*} k\left(\widehat{\gamma}(0,t)-\widehat{\gamma}(0,1)\right)^2 \mathrm{d}t + \int_{t^*}^{1} k\left(\widehat{\gamma}(0,t)-\widehat{\gamma}(0,1)\right)^2 \mathrm{d}t =: A_n + B_n.
\end{aligned}$$

By following the steps leading to (2.62) we get from (2.28) (on a suitable probability space)

$$\begin{aligned}
\frac{1}{\log(1/\widetilde{t}_n)}A_n &\overset{\text{a.s.}}{=} \frac{1}{\log(1/\widetilde{t}_n)}\int_{\widetilde{t}_n}^{t^*}\left(\frac{B_{\mathrm{pre}}(t)}{t} - B_{\mathrm{pre}}(1)\right)^2 \mathrm{d}t(1+o(1)) \\
&\overset{(2.71)}{=} \frac{1}{\log(1/\widetilde{t}_n)}\int_{\widetilde{t}_n}^{t^*}\left(\sigma_{\widehat{\gamma},\gamma_{\mathrm{pre}}}\frac{W(t)}{t} - B_{\mathrm{pre}}(1)\right)^2 \mathrm{d}t(1+o(1)) \\
&= \frac{1}{\log(1/\widetilde{t}_n)}\int_{\log(\widetilde{t}_n)}^{\log(t^*)}\left(\sigma_{\widehat{\gamma},\gamma_{\mathrm{pre}}}\frac{W(e^y)}{e^{y/2}} - e^{y/2}B_{\mathrm{pre}}(1)\right)^2 \mathrm{d}y(1+o(1)).
\end{aligned}$$

By slightly adapting the arguments in the proof of Theorem 2.2 this term converges in probability to $\sigma^2_{\widehat{\gamma},\gamma_{\mathrm{pre}}}$. Furthermore $B_n/\log(1/\widetilde{t}_n) = \mathcal{O}_{\mathrm{P}}(1)\log^{-1}(1/\widetilde{t}_n) = o_{\mathrm{P}}(1)$ by (2.28). The result follows.

For the consistency results in (c) and (d) combine (a) and (b) with the conclusion of Theorem 2.3 to deduce for (e.g.) Q_{seq}

$$\begin{aligned}
Q_{\text{seq}} &\geq \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \left\{ t^*(1-t^*)\sqrt{k}\left(\widehat{\gamma}\,(0,t^*) - \widehat{\gamma}\,(t^*,1)\right)\right\}^2 \\
&= \left(\frac{1}{\sigma^2_{\widehat{\gamma},\max(\gamma_{\text{pre}},\gamma_{\text{post}})}} + o_{\mathrm{P}}(1)\right) k \left\{ \underbrace{t^*(1-t^*)}_{>0} \left(\underbrace{\widehat{\gamma}\,(0,t^*)}_{\xrightarrow{\mathrm{P}} \gamma_{\text{pre}}} - \underbrace{\widehat{\gamma}\,(t^*,1)}_{\xrightarrow{\mathrm{P}} \gamma_{\text{post}}} \right) \right\}^2 \xrightarrow[(n\to\infty)]{\mathrm{P}} \infty.
\end{aligned}$$

Note that assumption $k = k_{\text{pre}} = k_{\text{post}}$ was needed to deduce via (the equivalents of) (2.16)

$$\widehat{\gamma}\,(0,t^*) \xrightarrow{\mathrm{P}} \gamma_{\text{pre}} \quad \text{and} \quad \widehat{\gamma}\,(t^*,1) \xrightarrow{\mathrm{P}} \gamma_{\text{post}},$$

where by definition of Q_{seq} both extreme value index estimators rely on the *same* sequence k.

For the inconsistency in (d) of the test based on Q_{rec} (the proof for $Q^{\leftarrow}_{\text{rec}}$ is similar and hence omitted) combine the result of Theorem 2.3 with the continuous mapping theorem and part (a) of Corollary 2.3 to deduce

$$Q_{\text{rec}} \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{1}{\sigma^2_{\widehat{\gamma},\gamma_{\text{pre}}}} \sup_{t\in[t_0,1-t_0]} \left\{B_{\text{pre}}(t) - tB_{\text{pre}}(1)\right\}^2,$$

whence $Q_{\text{rec}} = \mathcal{O}_{\mathrm{P}}(1)$, $n \to \infty$. $\square$

3 Testing for changes in (extreme) VaR

In this chapter we develop tests for a change in an unconditional small quantile (Value-at-Risk, VaR, in financial time series analysis) based on an estimator motivated by extreme value theory. This so called Weissman (1978) estimator allows tests to be applied for extreme VaR, where tests proposed so far in the literature mostly fail. Consistency is shown under local alternatives, where multiple breaks may occur. In view of applications we allow for weakly dependent observations. A simulation study shows that in finite samples our tests compare favorably in the tail region with extant tests based on order statistic estimators and also with tail index break tests. An empirical example using crude oil returns serves to illustrate the practical use of our tests.

3.1 Motivation

Structural changes in financial time series are well documented (see Andreou and Ghysels, 2009, for an overview). Specifically, tests for changes in proxies for risk have been discussed in the literature for some time. For example, Inclan and Tiao (1994) proposed a test for variance breaks and found evidence for them in stock return data. More recently, tests for changes in the tail index were introduced by Quintos *et al.* (2001), who detected shifts in the tail index of south east Asian stock index returns during the Asian financial crisis of 1997-1998. Yet, to cite Diebold *et al.* (2000): 'For example, in financial risk management, interest centers not on the tail index per se, but rather on the extreme quantiles and probabilities. The ability to withstand big hits, quantified by specific probabilities, translates directly into credit ratings, regulatory capital requirements, and so on.' Certainly, this comment not only applies to the tail index, but also to the variance.

A further advantage of quantile-based change point tests is that a rejection is easier to interpret. For, e.g., the tests in Quintos *et al.* (2001) were derived under the null of strict stationarity. The null distribution will likely be influenced by a mere variance break (or a scale break if variance is infinite). Of course, a mere scale change leaves the tail index unaffected and should thus be accounted for under the null of tail index constancy. Hence a (non-)rejection of the null may always be due to variance breaks that were not properly accounted for under the null. We note that variance breaks are a general problem in change point detection, which was only solved in very specific instances (e.g., Xu, 2015). Similarly, the variance break test in Inclan and Tiao (1994) does not account for mere breaks in the mean (which would, in analogy to the previous example, not affect the variance) under the null and consequently suffers from the same problem. This problem however does not arise for quantile-based tests, because there a mean or variance break is no longer covered under the null of quantile constancy.

To address the possibility of structural changes in extreme quantiles in real data, we propose a retrospective method to detect extreme quantile breaks in this chapter. To the best of our knowledge no such tests exist in the literature. In the context of financial return series small quantiles are often referred to as Value-at-Risk (VaR). VaR is one of the most widely used risk measures in the financial industry, not only because of regulatory requirements, most prominently the Basel Accords for banking supervision (cf., e.g., Longin, 2000, Sec. 4.1.4), but also because of its conceptual simplicity, ease of computation and backtesting. However, despite its prominence in practice, VaR suffers from the theoretical defect of not (generally) being subadditive, where subadditivity means, roughly speaking, that the risk of holding two assets should be smaller than the sum of the risk of holding each asset separately. Hence, though VaR satisfies the homogeneity, monotonicity and translation invariance property, it fails the subadditivity requirement for a risk measure to be 'coherent' in the sense of Artzner *et al.* (1999), whom we also refer to for discussion of the importance of subadditivity.

Daníelsson *et al.* (2013) recently showed that if, heuristically, asset returns are jointly regularly varying, then VaR is subadditive for sufficiently small probabilities. In their Monte Carlo simluations they find that in some cases VaR estimates using order statistics nonetheless lead to substantial violations of subadditivity. This prompts them to investigate the Weissman (1978) estimator, a (semi-parametric) quantile estimator based on extreme value theory (EVT), which leads to far fewer subadditivity violations in simulations. Because of this practical superiority we focus on the Weissman (1978) estimator to detect changes in extreme quantiles. We do so for a further reason: Since we will be testing for changes in *unconditional* quantiles, we take a long-term perspective. And in stress tests and worst-case scenario analysis, where the long-term viability of institutions is under inspection, typically the unconditional VaR is most important for extremely small probabilities. This may require extrapolating outside the range of available data, as in, say, the estimation of a once-in-1000 daily loss based on only two years of trading, corresponding to roughly 500 observations. It is this area where conventional quantile estimators turn out to be (mostly) useless, but EVT-based ones, like the Weissman (1978) estimator, still provide consistent estimates.

In focusing on (long-term) changes in unconditional quantiles using EVT methods we complement the literature on (short-term) changes in conditional quantiles enhanced by EVT methods, which aims to improve conditional (extreme) VaR forecasts. We refer to Chavez-Demoulin *et al.* (2014) for one of the latest additions to this strand of the literature and the references therein for an overview. We stress that both conditional and unconditional estimation have their distinct advantages and drawbacks: Employing unconditional methods for a one-day ahead forecast disregards information used in conditional forecasts, leading to a clustering of VaR violations (see Chavez-Demoulin *et al.*, 2014, Fig. 2) and thus a higher risk of bankruptcy. On the other hand, the use of conditional VaR forecasts for the long-term can lead to severe under- or overestimation of (long-term) unconditional VaR (in much the same way as one realization of a random variable cannot be relied upon to give a good indication of its mean). To see the analogy more clearly note that the conditional distribution of returns X_{t+1} (on some information set $\mathcal{F}_t$), say $\mathrm{P}_{X_{t+1}|\mathcal{F}_t}$, is generally not equal to the unconditional distribution $\mathrm{P}_{X_{t+1}}$, though related via $\mathbb{E}[\mathrm{P}_{X_{t+1}|\mathcal{F}_t}] = \mathrm{P}_{X_{t+1}}$. Hence, if $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ and X_t is stationary and ergodic, then $\mathrm{P}(X_{t+1} \in B|\mathcal{F}_t)$ is stationary and ergodic as well as a (measurable) function of $X_t, X_{t-1}, \ldots$ and by the ergodic theorem $1/n \sum_{t=1}^{n} \mathrm{P}(X_{t+1} \in B|\mathcal{F}_t) \to \mathrm{P}(X_0 \in B)$ almost surely for any Borel set B, i.e., the time average of conditional quantiles estimates the unconditional quantile. In light of this, if risk management is based mostly on day-to-day conditional forecasts and they happen to be quite low for some time, as, e.g., in the two year period before the financial crisis in 2007-8, institutions will be more vulnerable in times of stress, as evidenced by the Lehman bankruptcy. It is our aim in this chapter to propose structural change tests for (extreme) unconditional quantiles.

To emphasize the importance of unconditional extreme quantiles (and hence of structural break tests for them) Dacarogna *et al.* (2001, p. 144) argue 'that for practical purposes the hedge against extreme risk must be decided on the basis of the *unconditional* distribution. For a large portfolio, it would be impossible to find counterparties to hedge in very turbulent states of the market. Like in the case of earthquakes, hedging this type of risk needs to be planned far in advance.' To highlight further the practical importance of unconditional extreme quantiles we refer to the examples in Daníelsson and de Vries (1997).

Among the pitfalls of EVT pointed out in Diebold *et al.* (2000) is, first, that standard extreme quantile estimators are highly non-linear functions of tail index estimators. And if finite-sample approximations for the latter's distribution are poor, those of the extreme quantile estimators will likely be even poorer. We remedy this problem by taking logarithms of the extreme quantile estimators, such that those same estimators are *linear* functions of tail index estimators (see also Remark 3.4 below). Second, the focus of EVT is on i.i.d. data. As financial time series are well-known to exhibit, e.g., conditional heteroskedasticity (see Bollerslev, Chou and Kroner, 1992), we explicitly allow for dependent data. Obviously, although our focus is on financial applications, our results can also be applied in a wide range of disciplines where heavy tails are encountered, such as insurance, teletraffic engineering, etc.

The outline of this chapter is as follows. In Section 3.2 we state the limiting distribution of our test statistic under the null and consistency results. A simulation study is conducted in Section 3.3. We illustrate our procedure with an example concerning crude oil returns in Section 3.4 and sum up in Section 3.5. Proofs are relegated to the Section 3.6.

3.2 Main results

3.2.1 Preliminaries

Let $\{X_i\}_{i\in\mathbb{N}}$ denote a strictly stationary sequence of r.v.s. The Weissman (1978) estimator is motivated by EVT, which deals with (left or right) tail properties of X_i. The survivor function $\bar{F}(x) := \bar{F}_i(x) := P(X_i > x) = 1 - F(x)$ of X_i is regularly varying with parameter $-1/\gamma$ (where $\gamma > 0$ is called the extreme value index), written $\bar{F} \in RV_{-1/\gamma}$, i.e.,

$$\bar{F}(x) = x^{-1/\gamma} L(x), \qquad \text{where} \quad \lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1 \quad \forall\, t > 0. \tag{3.1}$$

This means that X_i has Pareto-type tails. Such tail behavior can be found in variables from many diverse fields (cf., e.g., Resnick, 2007). In terms of the $(1-1/t)$-quantile

$$U(t) := F^{\leftarrow}\left(1 - \frac{1}{t}\right),$$

(where $^{\leftarrow}$ denotes the left-continuous inverse) (3.1) is equivalent to

$$\lim_{x\to\infty} \frac{U(tx)}{U(x)} = t^{\gamma} \qquad \forall\, t > 0, \quad \text{i.e., } U \in RV_{\gamma}. \tag{3.2}$$

Furthermore, $\alpha = 1/\gamma > 0$ is called the tail index.

To state the basic testing problem suppose we observe a time series $X_1, \ldots, X_n$ with regularly varying survivor functions $\bar{F}_1, \ldots, \bar{F}_n$ and based on these observations have to decide whether or not a change in an extreme (right-tail) $(1-p)$-quantile has occurred, i.e., test the hypothesis

$$\begin{aligned} \mathcal{H}_0: \quad & U_1\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right) \qquad \text{vs.} \\ \mathcal{H}_1^{\lessgtr}: \quad & U_1\left(\frac{1}{p}\right) = \ldots = U_{\lfloor nt^*\rfloor}\left(\frac{1}{p}\right) \lessgtr U_{\lfloor nt^*\rfloor+1}\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right), \end{aligned} \tag{3.3}$$

for some $t^* \in [t_0, 1-t_0]$ with $t_0 \in (0, 1/2)$ arbitrarily small. Typical values of p in applications will lie between 0.001 and 0.05. For technical reasons we need to assume henceforth that $p = p_n \to 0$ as $n \to \infty$. We will show in simulations that the typical choices of p will lead to reasonable finite-sample approximations. In view of applications (see Section 3.4 below) we also need to allow for multiple breaks, so for, e.g., two breaks the alternative would read as

$$\mathcal{H}_2: \quad U_1\left(\frac{1}{p}\right) = \ldots = U_{\lfloor nt_1^* \rfloor}\left(\frac{1}{p}\right) \lesseqgtr U_{\lfloor nt_1^* \rfloor+1}\left(\frac{1}{p}\right) = \ldots = U_{\lfloor nt_2^* \rfloor}\left(\frac{1}{p}\right) \lesseqgtr$$
$$U_{\lfloor nt_2^* \rfloor+1}\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right), \quad t_2^* - t_1^* \geq t_0,\ t_1^*, t_2^* \in [t_0, 1-t_0].$$

Remark 3.1. (a) Without loss of generality we focus on the right tail in our development (as is customary in EVT), since by premultiplying the r.v.s with -1 we might as well consider the left tail.

(b) We remark that our alternative encompasses a wide range of parameter breaks, e.g., a break in the tail index and / or the variance and / or the mean can all cause a change in an extreme quantile. Hence, we provide an omnibus test for a whole range of tail-relevant parameter breaks. Further, the test results are easier to interpret than those of variance break tests in, e.g., Inclan and Tiao (1994), or tail index break tests in, e.g., Quintos *et al.* (2001) and Chapter 2, for the reasons spelled out in the motivation.

(c) Under certain conditions our test will be consistent even if $|U_1(1/p) - U_n(1/p)| \to 0$ as $n \to \infty$ under $\mathcal{H}_1^{\lesseqgtr}$, i.e., an asymptotically negligible change under the alternative. See Remark 3.5 (b) below.

The estimation of extreme quantiles is closely related to the estimation of the extreme value index γ. To see this for a sample $X_{\lfloor ns \rfloor+1}, \ldots, X_{\lfloor nt \rfloor}$ let $k = k_n \in \mathbb{N}$ be a sequence such that

$$k \underset{(n\to\infty)}{\longrightarrow} \infty \quad \text{and} \quad \frac{k}{n} \underset{(n\to\infty)}{\longrightarrow} 0,$$

and denote

$$X_k(s,t,y) := \Big(\lfloor ky(t-s) \rfloor + 1\Big)\text{-largest value of } X_{\lfloor ns \rfloor+1}, \ldots, X_{\lfloor nt \rfloor}.$$

Now use (3.2) (with $t = k/(np)$ and $x = n/k$) to approximate

$$U\left(\frac{1}{p}\right) \approx U\left(\frac{n}{k}\right)\left(\frac{np}{k}\right)^{-\gamma}$$

$$\approx X_k\left(s,t,1\right)\left(\frac{np}{k}\right)^{-\widehat{\gamma}(s,t)} =: \widehat{x}_p\left(s,t\right), \tag{3.4}$$

where $\widehat{\gamma}\left(s,t\right)$ $(0 \leq s < t \leq 1)$ is an estimator of γ based on the $\Big(\lfloor k\left(t-s\right)\rfloor + 1\Big)$ largest order statistics of $X_{\lfloor ns\rfloor+1}, \ldots, X_{\lfloor nt\rfloor}$ and $\widehat{x}_p$ is known as the Weissman (1978) estimator.

Remark 3.2. The Weissman (1978) estimator picks up changes in scale via $X_k\left(s,t,1\right)$ and changes in the tail index via $\widehat{\gamma}\left(s,t\right)$. Since changes in the tail index will usually be accompanied by changes in scale, it seems reasonable to expect a test based on $\widehat{x}_p$ to pick up a tail index break more often than a test based on $\widehat{\gamma}$ alone. This is corroborated in the simulations in Section 3.3.

Our test statistic for the testing problem (3.3) is given by

$$Q_1^{EQ} := \sup_{t\in[t_0,1-t_0]} \left\{ \frac{\left[t(1-t)\log\left(\frac{\widehat{x}_p(0,t)}{\widehat{x}_p(t,1)}\right)\right]^2}{\int_{t_0}^{t}\left[s\log\left(\frac{\widehat{x}_p(0,s)}{\widehat{x}_p(0,t)}\right)\right]^2 \mathrm{d}s + \int_{t}^{1-t_0}\left[(1-s)\log\left(\frac{\widehat{x}_p(s,1)}{\widehat{x}_p(t,1)}\right)\right]^2 \mathrm{d}s} \right\}, \tag{3.5}$$

where $t_0 \in (0, 1/2)$. The general form of the test statistic is inspired by the ideas of self-normalization in Shao and Zhang (2010). The purpose of the normalization in the denominator is, first, to obviate the need to estimate the asymptotic variance of (the suitably scaled) $\log\Big(\widehat{x}_p(0,1)/U\left(1/p\right)\Big)$. The second purpose is to account for the one-change point alternative under which a variance estimator (which we would have to divide the numerator of (3.5) by to obtain a pivotal test statistic) could potentially be large in finite samples. This would counteract the larger values in the numerator of (3.5) and thus may lead to less powerful tests and in some cases nonmonotonic power, i.e., decreasing power for more distant alternatives. The key observation regarding the normalization in the denominator is that for $t = t^*$ both integrals will (under suitable conditions) be $\mathcal{O}_{\mathrm{P}}(1)$ under the one-change point alternative $\mathcal{H}_1^{\lessgtr}$ in (3.3). Note further that the factor $t(1-t)$ was introduced so that Q_1^{EQ} is invariant under a time reversion of the sample. Thus, a reversed sample always provides the same evidence against the null as a non-reversed sample.

For the alternative $\mathcal{H}_2^{\lessgtr}$ we propose the test statistic Q_2^{EQ} defined as

$$\sup\left\{\frac{\left[\frac{t_1(t_2-t_1)}{t_2}\log\left(\frac{\widehat{x}_p(0,t_1)}{\widehat{x}_p(t_1,t_2)}\right)\right]^2}{\int_{t_0/2}^{t_1}\left[s\log\left(\frac{\widehat{x}_p(0,s)}{\widehat{x}_p(0,t_1)}\right)\right]^2\mathrm{d}s+\int_{t_1}^{t_2-t_0/2}\left[(t_2-s)\log\left(\frac{\widehat{x}_p(s,t_2)}{\widehat{x}_p(t_1,t_2)}\right)\right]^2\mathrm{d}s}\right.$$
$$\left.+\frac{\left[\frac{(t_2-t_1)(1-t_2)}{1-t_1}\log\left(\frac{\widehat{x}_p(t_1,t_2)}{\widehat{x}_p(t_2,1)}\right)\right]^2}{\int_{t_1+t_0/2}^{t_2}\left[(s-t_1)\log\left(\frac{\widehat{x}_p(t_1,s)}{\widehat{x}_p(t_1,t_2)}\right)\right]^2\mathrm{d}s+\int_{t_2}^{1-t_0/2}\left[(1-s)\log\left(\frac{\widehat{x}_p(s,1)}{\widehat{x}_p(t_2,1)}\right)\right]^2\mathrm{d}s}\right\},$$

where the supremum is taken over $t_2 > t_1$, such that for $t_0 \in (0, 1/3)$

$$t_0 \le t_1 < t_2 \le 1 - t_0, \quad t_2 - t_1 \ge t_0,$$

Similar comments as above apply. As for Q_1^{EQ}, the periods of stationarity between breaks should be $\lfloor nt_0 \rfloor$ for the test based on Q_2^{EQ} to be consistent. The extensions to more than two breakpoints should be clear.

In order to derive limiting distributions for the above test statistics under $\mathcal{H}_0$ we impose the following conditions:

(B1) $\{X_i\}_{i\in\mathbb{N}}$ is a strictly stationary β-mixing process with continuous marginals and mixing coefficients $\beta(\cdot)$, such that

$$\lim_{n\to\infty}\frac{n}{r_n}\beta(l_n)+\frac{r_n}{\sqrt{k}}\log^2(k)=0$$

for sequences $\{l_n\}_{n\in\mathbb{N}} \subset \mathbb{N}, \{r_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$ tending to ∞ with $l_n = o(r_n)$, $r_n = o(n)$.

(B2) There exists a function $r = r(x, y)$, such that for all $x, y \in [0, 1 + \varepsilon]$ $(\varepsilon > 0)$

$$\lim_{n\to\infty}\frac{n}{r_n k}\sum_{1\le i,j\le r_n}\operatorname{Cov}\left(I_{\left\{X_i>U\left(\frac{n}{kx}\right)\right\}}, I_{\left\{X_j>U\left(\frac{n}{ky}\right)\right\}}\right)=r(x,y).$$

(B3) For some constant $C > 0$

$$\frac{n}{r_n k}\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U\left(\frac{n}{ky}\right)<X_i\le U\left(\frac{n}{kx}\right)\right\}}\right]^4 \le C(y-x) \quad \forall\ 0\le x<y\le 1+\varepsilon,\ n\in\mathbb{N}.$$

(B4) There exist $\rho < 0$ and a function $A(\cdot)$ that is eventually positive or negative with $\lim_{t\to\infty} A(t) = 0$, s.t.

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \qquad \forall\, x > 0,$$

where $\sqrt{k}A(n/k) \to 0$ as $n \to \infty$.

(B5) $\lim_{n\to\infty} \frac{np}{k} = 0$, $\lim_{n\to\infty} k^{-1/2} \log{(np)} = 0$.

(B6) The sequence k satisfies

$$\frac{U\left(1/p\right)}{U\left(n/k\right)} \left(\frac{np}{k}\right)^\gamma - 1 = o\left(\frac{1}{\sqrt{k}}\right). \tag{3.6}$$

Remark 3.3. (a) For a discussion of the conditions **(B1)**-**(B4)** including their applicability we refer to Chapter 2. Note that conditions **(B1)**-**(B4)** are precisely conditions **(A1)**-**(A4)** in Chapter 2. Theorem 2.1 shows that under these conditions we have under a Skorohod construction that for some small $\widetilde{\delta} > 0$ and every $\nu \in [0, 1/2)$

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}-\widetilde{\delta}}} y^{\nu/\gamma} \left| \sqrt{k} \left(\frac{1}{k} \sum_{i=1}^{\lfloor nt \rfloor} I_{\left\{X_i > yU(n/k)\right\}} - y^{-1/\gamma} t \right) - W(t, y^{-1/\gamma}) \right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0 \tag{3.7}$$

holds. For the definition of $W(\cdot,\cdot)$ we refer to Theorem 2.1. This result will be crucial in the proof of Theorem 3.1 below. In Chapter 2 we used this result to derive change point tests for the tail index for the following estimators: the Hill (1975) estimator, the moments ratio estimator (see Daníelsson *et al.*, 1996) and the Csörgő and Viharos (1998) estimator; see Examples 2.3 - 2.5. Throughout this section we will consider $\widehat{x}_p$ based on theses tail index estimators.

(b) The assumption $\lim_{n\to\infty} \frac{np}{k} = 0$ in **(B5)** is reasonable, since if, e.g., $\lim_{n\to\infty} \frac{np}{k} = c \in (0, \infty)$, we are in a less extreme region, where non-parametric quantile estimators could be used.

The second assumption $\lim_{n\to\infty} k^{-1/2} \log{(np/k)} = \lim_{n\to\infty} k^{-1/2} \log{(np)} = 0$ in **(B5)** restricts p not to become too small too fast relative to n/k, i.e., describes the barrier beyond which extrapolation becomes unfeasible.

(c) If a d.f. F satisfies

$$1 - F\left(x\right) = Cx^{-\alpha} \left(1 + \mathcal{O}(x^{-\beta})\right), \qquad x \to \infty;\ C, \alpha, \beta > 0,$$

which is quite a general subclass of (3.1) (satisfied for, e.g., ARCH(1)-processes; see Drees (2003, p. 634)), then

$$U\left(1/y\right) = F^{-1}\left(1-y\right) = C^{1/\alpha} y^{-1/\alpha}\left(1+\mathcal{O}(y^{\beta/\alpha})\right), \quad y \downarrow 0. \tag{3.8}$$

Hence, in this case condition **(B6)** does not impose an extra restriction on the choice of k, since (3.6) can easily seen to be implied by (3.8).

3.2.2 Results under the null and the alternative

Theorem 3.1. *Suppose* **(C1)**-**(C6)** *hold. Then for* $\widehat{x}_p\left(\cdot,\cdot\right)$ *defined in* (3.4) *it holds under a Skorohod construction that for some* $\sigma_{\widehat{\gamma},\gamma} > 0$

$$\sup_{\substack{0\le t_1<t_2\le 1\\ t_2-t_1\ge t_0}} \left| \frac{\sqrt{k}}{\log(k/(np))} \log\left(\frac{\widehat{x}_p(t_1,t_2)}{U(1/p)}\right) - \sigma_{\widehat{\gamma},\gamma} \frac{W(t_2)-W(t_1)}{t_2-t_1} \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0,$$

where $t_0 \in (0,1)$ *and* $\left\{W\left(t\right)\right\}_{t\in[0,1]}$ *is distributed as a standard Brownian motion.*

Corollary 3.1. *Under the conditions of Theorem 3.1 the following convergences hold for* $t_0 \in (0,1/2)$ *for* Q_1^{EQ} *and* $t_0 \in (0,1/3)$ *for* Q_2^{EQ}*:*

$$Q_1^{EQ} \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}$$

$$\sup_{t\in[t_0,1-t_0]} \frac{\left\{W(t)-tW(1)\right\}^2}{\int_{t_0}^{t}\left[W(s)-\frac{s}{t}W(t)\right]^2 \mathrm{d}s + \int_{t}^{1-t_0}\left[W(1)-W(s)-\frac{1-s}{1-t}\left(W(1)-W(t)\right)\right]^2 \mathrm{d}s},$$

$$Q_2^{EQ} \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \sup_{\substack{t_0\le t_1<t_2\le 1-t_0\\ t_2-t_1\ge t_0}} \left\{ \frac{\left\{W(t_1)-\frac{t_1}{t_2}W(t_2)\right\}^2}{W_{t_0}^{(1)}(0,t_1)+W_{t_0}^{(2)}(t_1,t_2)} \right.$$
$$\left. + \frac{\left\{W(t_2)-\frac{1}{1-t_1}\left[(1-t_2)W(t_1)+(t_2-t_1)W(1)\right]\right\}^2}{W_{t_0}^{(1)}(t_1,t_2)+W_{t_0}^{(2)}(t_2,1)} \right\},$$

where $\left\{W\left(t\right)\right\}_{t\in[0,1]}$ *is distributed as a standard Brownian motion and*

$$W_{t_0}^{(1)}(t_1,t_2) := \int_{t_1+t_0/2}^{t_2}\left[W(s)-W(t_1)-\frac{s-t_1}{t_2-t_1}\left(W(t_2)-W(t_1)\right)\right]^2 \mathrm{d}s,$$

$$W_{t_0}^{(2)}(t_1, t_2) := \int_{t_1}^{t_2 - t_0/2} \left[W(t_2) - W(s) - \frac{t_2 - s}{t_2 - t_1} \left(W(t_2) - W(t_1) \right) \right]^2 \mathrm{d}s.$$

Remark 3.4. As a result of equation (3.16) below we could have based our change point tests equally well on the difference $\widehat{x}_p(0,t)/\widehat{x}_p(t,1) - 1$ instead of log-differences $\log\left(\widehat{x}_p(0,t)/\widehat{x}_p(t,1)\right)$ in Q_1^{EQ}. However, as reported in simulations in Gomes and Pestana (2007, Sec. 3.4), the asymptotic distribution of $\widehat{x}_p$ tends to be a worse fit to its finite-sample one than for $\log(\widehat{x}_p)$. Intuitively, this may be explained by the fact that $\log\left(\widehat{x}_p\right) = \log\left(X_k(0,1,1)\right) - \widehat{\gamma} \log\left(\frac{np}{k}\right)$ is a linear of function of $\widehat{\gamma}$, while $\widehat{x}_p = X_k(0,1,1)\left(\frac{np}{k}\right)^{-\widehat{\gamma}}$ is a non-linear function of $\widehat{\gamma}$ and it is precisely $\widehat{\gamma}$ upon which its asymptotic expansion in (3.16) below rests (see also Drees, 2003, p. 628).

Theorem 3.2. *Under $\mathcal{H}_1^{>}$ ($\mathcal{H}_1^{<}$) let the triangular array $\{X_{i,n}\}_{i=1,\ldots,n,\, n\in\mathbb{N}}$ be given by*

$$X_{i,n} := \begin{cases} Y_i^{\mathrm{pre}}, & i \in I_{\mathrm{pre}} := \left\{1, \ldots, \lfloor nt^* \rfloor\right\}, \\ Y_i^{\mathrm{post}}, & i \in I_{\mathrm{post}} := \left\{\lfloor nt^* \rfloor + 1, \ldots, n\right\}, \end{cases}$$

where $\{Y_i^{\mathrm{pre}}\}_{i\in\mathbb{N}}$ and $\{Y_i^{\mathrm{post}}\}_{i\in\mathbb{N}}$ both satisfy conditions **(B1)**-**(B6)** *with*

$$k,\ \gamma_{\mathrm{pre}},\ U_{\mathrm{pre}}(\cdot),\ r_{\mathrm{pre}}(\cdot,\cdot), \quad \text{and} \quad k,\ \gamma_{\mathrm{post}},\ U_{\mathrm{post}}(\cdot),\ r_{\mathrm{post}}(\cdot,\cdot) \tag{3.9}$$

respectively. If additionally

$$\left| \frac{\sqrt{k}}{\log\left(\frac{k}{np}\right)} \log\left(\frac{U_{\mathrm{post}}(1/p)}{U_{\mathrm{pre}}(1/p)}\right) \right| \underset{(n\to\infty)}{\longrightarrow} \infty \tag{3.10}$$

holds under $\mathcal{H}_1^{\lessgtr}$, then the test based on Q_1^{EQ} is consistent.

Remark 3.5. (a) For a discussion on the common choice of k in (3.9) we refer to Remark 2.5.

(b) In the sense that $U_{\mathrm{post}}(1/p)/U_{\mathrm{pre}}(1/p) \to 1$ may hold under (3.10) by **(B5)** (in which case it necessarily holds that $\gamma_{\mathrm{pre}} = \gamma_{\mathrm{post}} = \gamma$), we may say we have consistency under *local* alternatives. If under (say) $\mathcal{H}_1^{<}$ we have $k = n^{\xi}$ and $p = n^{-\nu}$ for $\xi > 1 - \nu$ and $0 < \nu < 1/(2\gamma + 1)$, then even $U_{\mathrm{post}}(1/p) - U_{\mathrm{pre}}(1/p) \to 0$ may hold (and thus we may detect asymptotically infinitesimal changes in the tail behavior). To see this, bound the left-hand side of (3.10) from below using $\log(x) > 1 - 1/x$ for $x > 1$ by

$$\frac{\sqrt{k}}{\log\left(k/(np)\right)} \frac{U_{\mathrm{post}}(1/p) - U_{\mathrm{pre}}(1/p)}{U_{\mathrm{post}}(1/p)}.$$

With the prescribed choices of ξ and ν and $U_{\text{post}}(1/p) \leq p^{-\gamma-\varepsilon}$ (cf., e.g., Bingham *et al.*, 1987, Prop. 1.3.6) for some $\varepsilon > 0$ and p sufficiently small, **(B5)** can easily seen to be satisfied and

$$\frac{\sqrt{k}}{\log\left(k/(np)\right) U_{\text{post}}(1/p)} \underset{(n\to\infty)}{\longrightarrow} \infty.$$

(c) Using arguments as in the proof of Theorem 3.2, a similar consistency result can be established for Q_2^{EQ} under $\mathcal{H}_2^{\lessgtr}$ and (3.10) for at least one of the two breakpoints. The analogy should be sufficiently clear. Note that the periods of stationarity between the breaks must be at least $\lfloor nt_0 \rfloor$ observations long for Q_2^{EQ} to detect them.

3.3 Simulations

All simulations were run using *R version 3.2.3* (R Core Team, 2015). As mentioned in the motivation our tests will most likely be employed in contexts where small probabilities are of interest. Hence we focus on $p = 0.005, 0.01, 0.05, 0.1$. Note that for each p we test a different null hypothesis. We investigate the test statistic Q_1^{EQ} for these *fixed* choices of p. Recall that our asymptotic results required $p = p_n \to 0$. Yet it will turn out that for these small choices of p the asymptotic distribution is a good approximation to that of Q_1^{EQ}. For purposes of comparison we also use the change point test proposed in Shao and Zhang (2010). We call the appertaining test statistic Q_1^{SZ} for short, where the only difference with Q_1^{EQ} is that $\widehat{x}_p(t_1, t_2)$ is replaced with the $(\lfloor n(t_2 - t_1)p \rfloor + 1)$-largest value of $X_{\lfloor nt_1 \rfloor + 1}, \ldots, X_{\lfloor nt_2 \rfloor}$, i.e., a mere order-statistic estimator of the $(1-p)$-quantile. To our knowledge Shao and Zhang's (2010) crucial Assumption 3.2 needed for convergence of the test statistic, has not been checked for the quantile functional and a particular time series model. We also compare our test with a tail index break test (again based on the same testing functional, say Q_1^{TI}, where the only difference with Q_1^{EQ} is that $\widehat{x}_p(t_1, t_2)$ is replaced with $\widehat{\gamma}(t_1, t_2)$), to investigate if there are any gains in power, as conjectured in Remark 3.2.

Quantile	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.99
Q_1^{EQ}	11.57	15.95	22.92	33.77	55.88	80.21	106.9	147.8
Q_2^{EQ}	111.7	134.8	165.4	203.6	287.0	367.4	463.9	592.5

Table 3.1: Critical values for Q_1^{EQ} and Q_2^{EQ} test with $t_0 = 0.2$

We first simulate strictly stationary $\{|X_i|\}$ from the MA(2)-model

$$X_i = Z_i + 0.9 \cdot Z_{i-1} + 0.7 \cdot Z_{i-2}, \tag{MA2}$$

with i.i.d. t_3-distributed innovations Z_i, where t_ν denotes a Student's t-distribution with ν degress of freedom. Recall that the t_ν-distribution has tail index $\alpha = \nu$. So X_1 is also heavy tailed with tail index $\alpha = 3$, since by Lemma 5.2 of Datta and McCormick (1998), as $x \to \infty$,

$$\frac{1 - F(x)}{1 - F_Z(x)} \to \frac{2.6}{2} = \frac{1 + 0.9 + 0.7}{2}.$$

Furthermore, we consider a strictly stationary ARCH(1)-process $\{X_i\}$ generated according to

$$X_i = \left(\alpha_0 + \alpha_1 X_{i-1}^2\right)^{1/2} Z_i, \quad i \in \mathbb{N}, \tag{ARCH1}$$

where $\alpha_0, \alpha_1 > 0$ and $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. In applications, α_1 is typically less than but close to 1, whence in our simulations we use $\alpha_1 = 0.8$ and $\alpha_0 = 0.01$. The tail index α in this model can be calculated from the moment equation $\mathrm{E}(\alpha_1 Z_1^2)^{\alpha/2} = 1$, which results in $\alpha = 2.68$ for our choice of $\alpha_1 = 0.8$ (cf. Davis and Mikosch, 1998, Table 1). Its unconditional variance is given by $\mathrm{Var}(X_1) = \alpha_0/(1 - \alpha_1)$ if $\alpha_1 < 1$.

For the verification of **(B1)**-**(B4)** for these models we refer to Drees (2003, Sections 3.1 & 3.2), see also Examples 2.1 and 2.2. As for the 'verification' of **(B5)** we note that p is chosen small but not too small relative to the magnitude of n/k. We note the choice of k is a notoriously difficult issue in extreme value theory. Apart from the heuristic proposal in Drees (2003, p. 645) we are not aware of any guidelines in the case of dependent data, let alone data with possible structural breaks. Hence, we investigate how robust the asymptotics are with respect to different choices of k. Finally, condition **(B6)** is checked via (3.8) (cf. Remark 3.3).

We explore the behavior of our tests for $t_0 = 0.2$ and sample sizes of $n = 500, 2000$; values which were also chosen in Quintos *et al.* (2001). Throughout we use 5000 replications. Critical values for the test statistic Q_1^{EQ} are given in Table 3.1. We remark that the critical values for the Q_1^{TI}- and the Q_1^{SZ}-test are the same as for Q_1^{EQ}. We use the arguably most popular tail index estimator, the Hill (1975) estimator, in all subsequent simulations.

We first investigate size for the MA(2)-model. Table 3.2 shows that the Q_1^{EQ}- and the Q_1^{TI}-test are close to the nominal level for $n = 500$ and even closer for $n = 2000$. For most tests there is a only a slight tendency for rejections to decrease in k, which is encouraging given the sometimes high sensitivity of extreme quantile estimators on k (cf. Drees, 2003, Sec. 5). The results for the ARCH(1)-process are largely similar to those for the MA(2)-process except that rejections are more frequent, for $n = 500$

Model	Test	p	$n = 500$			$n = 2000$		
			k/n					
			0.08	0.16	0.24	0.08	0.16	0.24
(MA2)	Q_1^{EQ}	0.005	7.7	5.4	4.2	5.9	5.6	4.8
		0.01	8.8	5.3	4.5	6.4	5.3	4.8
		0.05	6.4	6.4	6.1	5.5	5.4	5.2
		0.1	8.0	5.8	6.3	6.7	5.3	5.1
	Q_1^{TI}		4.3	3.9	3.2	4.3	4.0	4.2
(ARCH1)	Q_1^{EQ}	0.005	11	9.3	8.2	9.2	7.9	7.0
		0.01	12	9.4	7.8	9.1	7.9	7.5
		0.05	9.3	9.3	9.0	6.6	6.8	6.9
		0.1	9.2	7.6	7.9	6.5	6.3	6.5
	Q_1^{TI}		3.9	4.7	4.5	6.7	6.5	5.7

Table 3.2: Empirical sizes in % at 5%-level for trajectories of length n according to (MA2) and (ARCH1)

	$n = 500$				$n = 2000$			
p	0.005	0.01	0.05	0.1	0.005	0.01	0.05	0.1
(MA2)	28	20	9.8	7.4	16	11	7.1	6.0
(ARCH1)	25	21	13	10	18	14	8.2	7.0

Table 3.3: Empirical sizes in % at 5%-level for Q_1^{SZ} test

more so than for the larger sample size. For $n = 2000$ however, the results are quite satisfactory, particularly compared with some other applications of extreme quantile estimation as in Drees (2003, Figs. 6 & 7). We conjecture that this is due to the self-normalization, which was already reported in Shao and Zhang (2010) to lead to improved size, at the cost though of slightly lower power.

Table 3.3 shows that the Q_1^{SZ}-test essentially yields acceptable results only for combinations $n = 2000$, $p = 0.05, 0.1$ and $n = 500$, $p = 0.1$. For the remaining p and n combinations its size distortion can be quite dramatic, which is why comparisons with Q_1^{EQ} will only be fair for the mentioned combinations.

We will focus on two ways in which $\mathcal{H}_1^{\lessgtr}$ in (3.3) can be true: a change in the variance and the tail index. Of course a tail index change is usually accompanied by a variance change. E.g. in the case of an ARCH(1)-process this can be seen by noting from the above formulas that increasing α_1 while leaving α_0 constant will lead to an increase in the variance and the tail index at the same time.

Consider the MA(2)-model called $\mathrm{MA}_{\mathcal{H}_1}$ with a variance break in the error distribution, specifically, $X_{i,n}^{break} = Z_{i,n} + 0.9Z_{i-1,n} + 0.7Z_{i-2,n}$, where

$$Z_{i,n} \overset{\text{i.i.d.}}{\sim} \begin{cases} t_3, & i \leq \lfloor nt^* \rfloor, \\ 0.5 \cdot t_3, & i > \lfloor nt^* \rfloor. \end{cases}$$

Note that the break in the variance of $X_{n,i}$ does not occur from one period to the next, but rather takes three periods to take full effect. Clearly, rejections in Table 3.4 decrease as the breakpoint t^* moves further away from the middle of the observations. Moreover, as might intuitively be expected, for smaller values of p (i.e., for more extreme quantiles) it becomes a bit harder for the tests to reject the null if $t^* = 0.5$ and a lot harder if $t^* = 0.75$. (By symmetry of the test statistics, similar conclusions would hold if $t^* = 0.25$.) As n becomes larger we see a marked increase in rejections. Note that we have simulated under the null of tail index constancy (a mere variance change leaves the tail index unaffected), and accordingly the test based on Q_1^{TI} should not reject any more than in Table 3.2. In fact, it even rejects less. This is somewhat surprising, since one would have expected the test to have mistaken the variance break for a tail index break and hence reject more than under (MA2). The performance of the Q_1^{SZ}-test is not satisfactory. In Table 3.5 for $p = 0.005$ and $t^* = 0.75$ it rejects less frequently than under the null for $n = 500, 2000$. For $p = 0.05, 0.1$ its performance is worse than that of Q_1^{EQ} for all t^*, while being more oversized than Q_1^{EQ} under the null.

Results for two ARCH(1)-processes $\{X_{i,n}^{\text{break}}\}$ where parameter breaks occurred are also shown in Table 3.4. Here,

$$X_{i,n}^{\text{break}} = \begin{cases} \left(\alpha_{0,\text{pre}} + \alpha_{1,\text{pre}} X_{i-1}^2\right)^{1/2} Z_i, & i \leq \lfloor nt^* \rfloor, \\ \left(\alpha_{0,\text{post}} + \alpha_{1,\text{post}} X_{i-1}^2\right)^{1/2} Z_i, & i > \lfloor nt^* \rfloor, \end{cases}$$

with $\alpha_{0,\text{pre}} = \alpha_{0,\text{post}} = 0.01$ and $\alpha_{1,\text{pre}} = 0.8 > 0.3 = \alpha_{1,\text{post}}$, where the resulting process is referred to as $\mathrm{ARCH}_{\mathcal{H}_{1a}}$, and with $\alpha_{0,\text{pre}} = 0.01 > 0.0211 = \alpha_{0,\text{post}}$ and again $\alpha_{1,\text{pre}} = 0.8 > 0.3 = \alpha_{1,\text{post}}$, where this process will be termed $\mathrm{ARCH}_{\mathcal{H}_{1b}}$. Both processes have a break in the tail index from $\alpha_{\text{pre}} = 2.68$ to $\alpha_{\text{post}} = 8.36$. But in the case of $\mathrm{ARCH}_{\mathcal{H}_{1b}}$ the parameter $\alpha_{0,\text{post}}$ was chosen such that the pre- and post-break 0.9-quantile are the same (such that strictly speaking we are simulating under the null of constancy of the 0.9-quantile), but the higher quantiles are larger in the pre-break period, due to the larger tail index (e.g., the 0.995-quantile is 0.95 pre-break and 0.54 post-break).

The picture for Q_1^{EQ} under $\mathrm{ARCH}_{\mathcal{H}_{1a}}$ is roughly the same as in the MA case: rejections increase markedly in n and the proximity of the breakpoint to the middle.

Model	Test	p	t^*	$n = 500$			$n = 2000$		
				k/n					
				0.08	0.16	0.24	0.08	0.16	0.24
$\mathrm{MA}_{\mathcal{H}_1}$	Q_1^{EQ}	0.005	0.5	31	27	26	62	69	74
			0.75	10	6.7	6.1	21	25	28
		0.01	0.5	42	39	36	79	84	86
			0.75	15	11	10	34	39	42
		0.05	0.5	75	77	77	99	100	100
			0.75	37	37	37	81	82	82
		0.1	0.5	73	88	90	100	100	100
			0.75	42	52	54	86	93	93
	Q_1^{TI}		0.5	2.5	1.9	1.6	2.7	2.1	1.6
			0.75	1.9	1.1	1.2	1.7	1.7	1.6
$\mathrm{ARCH}_{\mathcal{H}_{1a}}$	Q_1^{EQ}	0.005	0.5	37	34	31	75	78	79
			0.75	12	9.3	8.8	22	25	24
		0.01	0.5	46	40	35	81	82	83
			0.75	15	12	9.8	25	27	28
		0.05	0.5	47	51	50	89	90	90
			0.75	15	16	16	35	36	36
		0.1	0.5	38	49	51	82	92	92
			0.75	17	15	16	39	42	43
	Q_1^{TI}		0.5	5.7	10	11	39	48	49
			0.75	1.6	2.9	3.3	5.1	9.0	9.5
$\mathrm{ARCH}_{\mathcal{H}_{1b}}$	Q_1^{EQ}	0.005	0.5	19	16	14	45	44	39
			0.75	10	9.1	7.7	14	14	13
		0.01	0.5	20	17	14	43	40	35
			0.75	12	9.2	7.3	15	14	13
		0.05	0.5	13	13	12	19	20	19
			0.75	8.3	8.0	7.8	8.4	8.3	7.9
		0.1	0.5	13	9.5	9.8	10	7.5	7.6
			0.75	9.5	6.4	6.3	7.5	4.7	4.8
	Q_1^{TI}		0.5	9.1	14	13	45	54	53
			0.75	4.2	6.0	5.6	12	16	15

Table 3.4: Rejection rates in % at 5% level of tests for MA(2) with break in innovation distribution and ARCH(1) with parameter break

Model	t^*	$n = 500$				$n = 2000$			
		p							
		0.005	0.01	0.05	0.1	0.005	0.01	0.05	0.1
MA(2)	0.5	49	39	59	75	40	54	96	99
	0.75	24	18	20	29	15	17	56	75
$\text{ARCH}_{\mathcal{H}_{1a}}$	0.5	43	43	40	40	54	59	75	82
	0.75	19	18	15	14	16	15	23	28
$\text{ARCH}_{\mathcal{H}_{1b}}$	0.5	29	26	17	13	38	36	19	9.3
	0.75	22	20	17	12	16	15	11	7.3

Table 3.5: Rejection rates in % at 5%-level for Q_1^{SZ} test

The performance of Q_1^{TI} is dismal. For $t^* = 0.75$ it hardly rejects more than under the null (even less for $n = 500$) and for $t^* = 0.5$ its rejection frequency is markedly lower than that of Q_1^{EQ}. This may be explained as in Remark 3.2, because we do not only observe a tail index increase from 2.68 to 8.36 (leading to thinner tails), but also a variance decrease from 0.05 to 0.014. The Q_1^{SZ}-test always leads to fewer rejections than the one based on Q_1^{EQ}, sometimes rejections are less frequent by 25 percentage points.

Turning to $\text{ARCH}_{\mathcal{H}_{1b}}$ now, we see quite similar results for Q_1^{TI} as under $\text{ARCH}_{\mathcal{H}_{1a}}$, while those for $Q - 1^{EQ}$ are expectedly different. Rejections for $p = 0.1$ (where the 0.9-quantile was the same pre- and post-break) are only slightly higher than under the null, which demonstrates some robustness of our procedure in the absence of stationary marginal distributions that nonetheless share the same 0.9-quantile. For smaller values of p rejections now increase such that for $p = 0.005$ they rival that of Q_1^{TI} for $n = 2000$ and particularly for $n = 500$. $\text{ARCH}_{\mathcal{H}_{1b}}$ is the only model in our simulation study where Q_1^{SZ} is on par with $Q - 1^{EQ}$ in terms of rejections.

Summing up the simulation results, it seems that Q_1^{EQ} leads to good size and power properties for any choice of k. Q_1^{SZ} is lagging Q_1^{EQ} in almost all cases where they have comparable size under the null. Furthermore, as speculated in Remark 3.2, tests based on quantile estimators will usually offer better performance than those based on tail index estimators under the common alternative. More specifically, if a quantile change only happens very far out in the tails (roughly $p \leq 0.01$), only Q_1^{TI} and Q_1^{EQ} will be applicable, where the latter is clearly preferred when, as will generally be the case, a tail index and a variance change occur simultaneously.

3.4 Application

We now apply our test to a time series of crude oil log-returns. The application of EVT methods for VaR calculation has met with success in energy markets, see Marimoutou *et al.* (2009) and the references therein. However, shifts in volatility of oil markets can plausibly occur, as oil prices are influenced by, for instance, geopolitics (e.g., wars in the conflict-prone middle east) or OPEC actions. Hence, to guard against the risk of (mis)estimating VaR based on a sample with different VaRs, leading to biased results, one should apply a structural change test like ours first. Note that a bias in both directions is undesirable: If VaR is overestimated, too much capital (that could be used more productively elsewhere) will be put aside as a buffer against extreme price swings. Underestimation of VaR leads to lower capital reserves and thus a higher risk of bankruptcy.

Specifically, we work with $n = 2040$ West Texas Intermediate (WTI) log-returns from 1987 to 1994 downloaded from *http://research.stlouisfed.org/fred2/*. The time series is displayed in the top part of Figure 3.1. It covers the time period of the Iraqi invasion in Kuwait in August of 1990, which led to tensions on the oil market. We test whether this shock has led to a change in an extreme quantile of the log-losses (i.e., the relevant VaR for an investor who is invested in oil). We might as well have analyzed the right tail, which would have led to similar conclusions. From simply eyeballing the return series one may suspect an ARCH-type volatility cluster around 1991. However, knowing about the invasion one may also discern a relatively calm regime up until August of 1990 (the beginning of the invasion), followed by a very volatile period through to February of 1991 (coinciding roughly with the end of the First Gulf War) and a very calm regime for the rest of the sample. This would speak in favor of different regimes with their respective different unconditional VaRs. See Stărică and Granger (2005) for modeling approaches in this direction.

Before applying our test, we check that the time series is indeed heavy-tailed. To that end consider Figure 3.2, where estimates of the tail index of the WTI log-losses are plotted against the number of order statistics k used in the estimation. Panel (a) shows Hill and moments estimates for the original data, where the moments estimator can be used for general $\gamma \in \mathbb{R}$ (cf. de Haan and Ferreira, 2006, Section 3.5). Both plots are exclusively positive. However, the Hill estimates in particular heavily depend on the choice of k and it is difficult to read off a concrete estimate from a stable portion of the graph. A more stable plot can often be obtained by an appropriate shift of the data (recall that the tail index is invariant under location shifts). Panel (b) in Figure 3.2 displays the estimates for the data with the minimum log-loss subtracted from each observation, so that all observations are non-negative. Now both plots are smoother and the Hill estimates stabilize around a value of 0.07 for $k \geq 300$. The shaded area around the Hill estimates signifies 95%-confidence intervals calculated using the variance estimator from Theorem 2.2 (a), so that we can reject the null

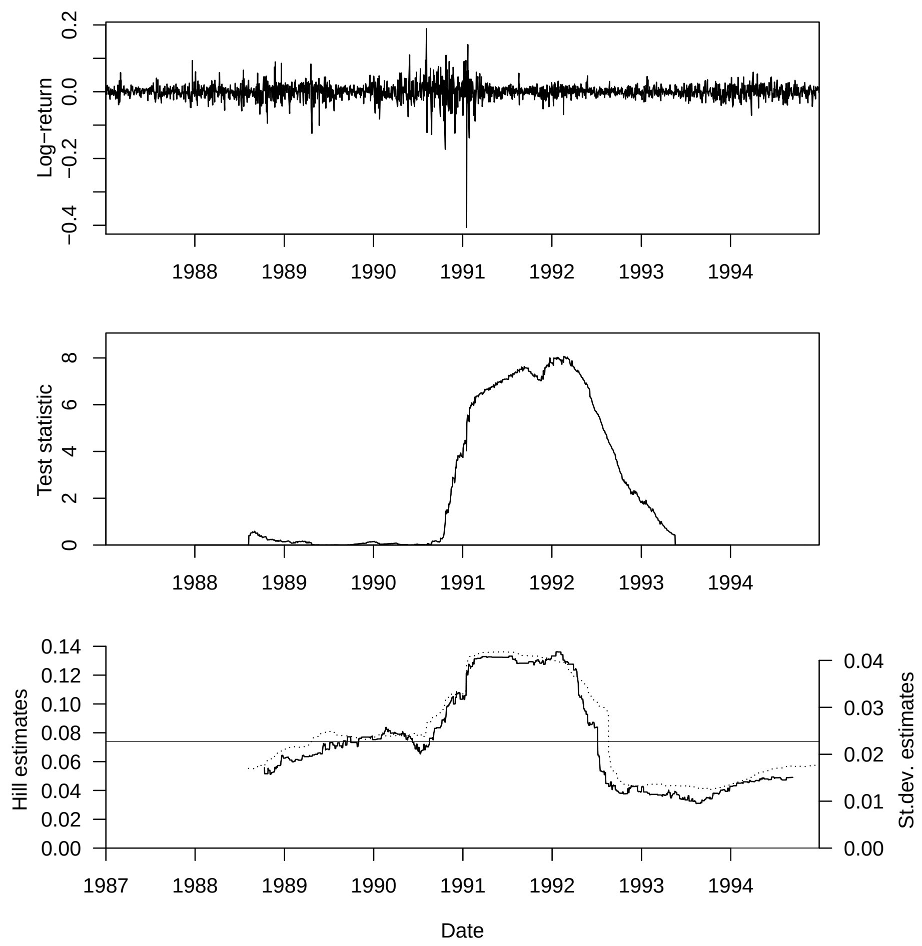

Figure 3.1: Top: Time series of WTI log-returns. Middle: Test statistics for shifted WTI log-losses. Bottom: Rolling window Hill (solid) and standard deviation (dotted) estimates of shifted log-losses. Solid horizontal line indicates Hill estimate based on whole sample.

$\gamma = 0$ for all values of k.

Having verified that the (positive) log-losses are indeed heavy-tailed, we now test for a breakpoint in the $(1-p)$-quantile of the shifted log-losses. However, doing so for different choices of p of course introduces a multiple testing problem. To avoid this, consider what we call the *quantile stability plot* in Figure 3.3. There, rolling window quantile estimates based on subsamples of length $\lfloor nt_0 \rfloor$ ($t_0 = 0.2$) are plotted

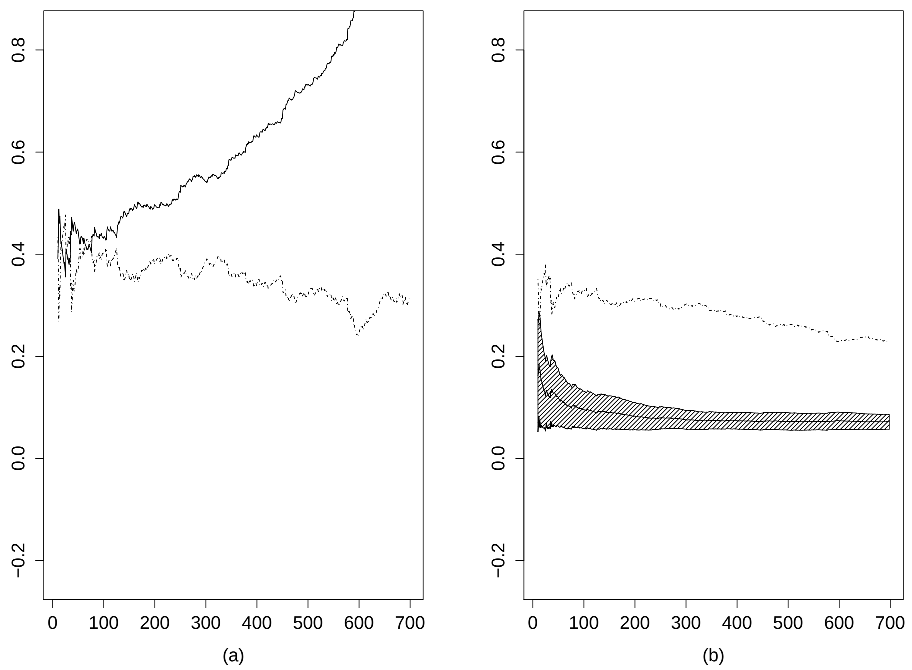

Figure 3.2: Hill (solid) and moment (dot-dashed) estimates of (a) WTI log-losses and (b) shifted WTI log-losses as a function of k

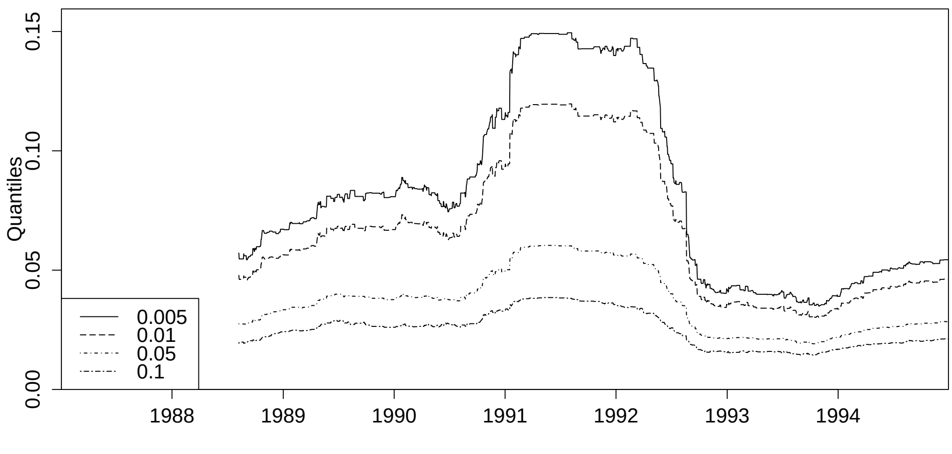

Figure 3.3: Rolling window estimates of $(1-p)$-quantiles ($p = 0.1, 0.05, 0.01, 0.005$) for shifted WTI log-losses

for different choices of p. (We used $k/n = 0.16$ and shifted log-losses to calculate the Hill estimator.) In view of the simulation results, we would choose a large value of p if all quantiles were to change by roughly the same order (which is the behavior expected under a variance break), and a small value of p if larger quantiles seem to change more than smaller ones (which would roughly correspond to a tail index break). The quantile estimates for $p = 0.005$ change more (they more than triple from lowest to highest) than those for $p = 0.1$ (they double). However, since the $(1 - 0.1)$-quantile can be estimated much more precisely, we choose to test for a change in the $(1 - 0.1)$-quantile.

In view of our prior knowledge about the Iraqi invasion it seems acceptable to use $t_0 = 0.2$ (as was done in the simulation section), which translates into knowing that the break (if any) occured between July of 1988 and May of 1993. The results are displayed in the middle of Figure 3.1 for $k/n = 0.16$ (i.e., $k = 326$) and the Hill estimator. The supremum is attained in the beginning of 1991, around the end of the First Gulf War. Yet it is far from the 5%-critical value of 80.21, such that a rejection is not warranted.

To investigate the possibility that the non-rejection is due to the presence of possible three breakpoints (under which consistency of the test based on Q_1^{EQ} has not been established) consider the bottom part of Figure 3.1. There, rolling window estimates based on subsamples of the shifted log-losses of length of $\lfloor nt_0 \rfloor$ for $t_0 = 0.2$ are plotted for the tail index and the standard deviation. This plot also points to the possible presence of three structural breaks. Remarkably, the estimates of the two parameters seem to move in lockstep, even though the underlying parameters they estimate could in principle evolve quite differently. However, since both estimates are affected by large outliers, such a behavior is not all that surprising after all. From eyeballing the series and the rolling window estimates one may suspect that the supremum of Q_2^{EQ} is attained for values around $t_1 = 0.4$ and $t_2 = 0.55$. Taking these as starting values, a local maximum is indeed reached for $t_1 = 0.45$ and $t_2 = 0.57$, which gives a lower bound of 56 for the value of Q_2^{EQ}. This is of course far from any sensible critical region, see Table 3.1. A precise evaluation of Q_2^{EQ} is computationally infeasible, because of the involved sorting required for calculating $\widehat{x}_p(t_1, t_2)$. It is nevertheless safe to say that we did not find solid evidence for (unconditional) extreme quantile breaks in the WTI log-loss series. The results are quite similar for different values of k and p and are not reported here.

As a final plausibility check of our result we fit a benchmark GARCH(1,1)-model with t_ν-noise to the WTI log-losses. Indeed, by applying standard statistical tests to the (raw and squared) standardized residuals, we find no evidence for any autocorrelation. Thus, the (stationary) GARCH specification seems to capture adequately the type of (apparently conditional) heteroskedasticity observed in the WTI log-losses.

3.5 Conclusion

We propose change point tests for (extreme) quantiles, where the focus on the extremes is novel. Our alternative covers a wide range of breaks in different parameters determining tail behavior. We thus offer an omnibus test relevant to (financial) risk management whose results have a clearer interpretation than extant test (e.g., in Quintos *et al.*, 2001; Inclan and Tiao, 1994). Our tests are based on the Weissman (1978) estimator, which has several advantages over mere order statistic based ones: it produces fewer subadditivity violations, has lower variance and consistently estimates extreme quantiles. The first advantage renders it the estimator of choice in practice. As demonstrated in the simulations, in a change point context the latter two properties directly translate into, first, higher power under the alternative in comparison with the test in Shao and Zhang (2010) based on an order statistic estimator, while, second, delivering reliable size under the null even for very small quantiles. Furthermore, our test provides reliably higher power in detecting tail-relevant changes than tests based on tail index estimators if, as is usually the case, a tail index break will be accompanied by a variance break. In an empirical application to WTI oil returns we find no evidence for extreme quantile breaks during the First Gulf War.

3.6 Proofs

Proof of Theorem 3.1: The first part of the proof resembles that of Theorem 3 in Einmahl *et al.* (2016). From (3.7) we may derive under a Skorohod construction that for $\nu \in [0, 1/2)$

$$\sup_{\substack{0\le t_1<t_2\le 1\\ t_2-t_1\ge t_0}} \sup_{y\ge y_0^{\gamma}-\widetilde{\delta}} y^{\nu/\gamma} \left| \sqrt{k}\left[\frac{1}{k(t_2-t_1)}\sum_{i=\lfloor nt_1\rfloor+1}^{\lfloor nt_2\rfloor} I_{\left\{X_i>yU(n/k)\right\}} - y^{-1/\gamma}\right] - \frac{W(t_2,y^{-1/\gamma})-W(t_1,y^{-1/\gamma})}{t_2-t_1}\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0, \quad (3.11)$$

where $W(t, y)$ is a continuous zero-mean Gaussian process with covariance function

$$\mathrm{Cov}(W(t_1, y_1), W(t_2, y_2)) = \min(t_1, t_2) r(y_1, y_2), \quad (3.12)$$

where $r(\cdot,\cdot)$ is defined in **(B2)**. Now use Einmahl *et al.* (2010, Lemma 5) similarly as in the proof of Einmahl *et al.* (2016, Theorem 3) to derive

$$\sup_{\substack{0\le t_1<t_2\le 1\\ t_2-t_1\ge t_0}}\left|\sqrt{k}\left[\left(\frac{X_k(t_1,t_2,1)}{U(n/k)}\right)^{-1/\gamma}-1\right]+\frac{W(t_2,1)-W(t_1,1)}{t_2-t_1}\right|\xrightarrow[(n\to\infty)]{\text{a.s.}}0. \tag{3.13}$$

As in the proof of Corollary 2.1 it then follows from (3.13) that $U(n/k)$ in (3.11) can be replaced by its empirical counterpart $X_k(t_1,t_2,1)$ when suitably accounting for it in the limit process:

$$\sup_{\substack{0\le t_1<t_2\le 1\\ t_2-t_1\ge t_0}}\sup_{y\ge y_0^\gamma-\widetilde{\delta}} y^{\nu/\gamma}\Bigg|\sqrt{k}\left[\frac{1}{\lfloor k(t_2-t_1)\rfloor}\sum_{i=\lfloor nt_1\rfloor+1}^{\lfloor nt_2\rfloor} I_{\{X_i>yX_k(t_1,t_2,1)\}}-y^{-1/\gamma}\right]$$
$$-\frac{W(t_2,y^{-1/\gamma})-W(t_1,y^{-1/\gamma})-y^{-1/\gamma}\left[W(t_2,1)-W(t_1,1)\right]}{t_2-t_1}\Bigg|\xrightarrow[(n\to\infty)]{\text{a.s.}}0.$$

The Hill estimator can be written as (see also Example 2.4)

$$\widehat{\gamma}(t_1,t_2)=\int_1^\infty\frac{1}{\lfloor k(t_2-t_1)\rfloor}\sum_{i=\lfloor nt_1\rfloor+1}^{\lfloor nt_2\rfloor} I_{\{X_i>yX_k(t_1,t_2,1)\}}\frac{\mathrm{d}y}{y},$$

whence

$$\sup_{\substack{0\le t_1<t_2\le 1\\ t_2-t_1\ge t_0}}\Bigg|\sqrt{k}\left[\widehat{\gamma}(t_1,t_2)-\gamma\right]$$
$$-\frac{\gamma}{t_2-t_1}\left[\int_0^1\{W(t_2,u)-W(t_1,u)\}\frac{\mathrm{d}u}{u}-\{W(t_2,1)-W(t_1,1)\}\right]\Bigg|\xrightarrow[(n\to\infty)]{\text{a.s.}}0. \tag{3.14}$$

We can see that $W(t):=\gamma\int_0^1 W(t,u)\frac{\mathrm{d}u}{u}-W(t,1)$ is distributed as a Brownian motion multiplied by $\sigma_{\widehat{\gamma},\gamma}$ defined in (2.19) by calculating the covariance function using (3.12). Then write (3.14) as

$$\sup_{\substack{0\le t_1<t_2\le 1\\ t_2-t_1\ge t_0}}\left|\sqrt{k}\left[\widehat{\gamma}(t_1,t_2)-\gamma\right]-\sigma_{\widehat{\gamma},\gamma}\frac{W(t_2)-W(t_1)}{t_2-t_1}\right|\xrightarrow[(n\to\infty)]{\text{a.s.}}0. \tag{3.15}$$

Identical expressions also hold for the other estimators from Remark 3.3 (a), where the only difference is that $\sigma_{\widehat{\gamma},\gamma}$ are now defined as in (2.18) and (2.21).

The rest of the proof generalizes that of Theorem 2.2 in Drees (2003). In view of (3.15) and $\log(x) \sim x-1$, $x \to 1$, we only need to verify

$$\frac{1}{\log\left(\frac{k}{np}\right)}\left(\frac{\widehat{x}_p(t_1,t_2)}{U(1/p)}-1\right) = \gamma - \widehat{\gamma}(t_1,t_2) + o_{\mathrm{P}}\left(1/\sqrt{k}\right), \tag{3.16}$$

where here and in the following all o- and $\mathcal{O}$-terms are uniform on $0 \le t_1 < t_2 \le 1$, $t_2 - t_1 \ge t_0$.

$$\begin{aligned}
\widehat{x}_p(t_1,t_2) - U(1/p) &= X_k(t_1,t_2,1)\left(\frac{np}{k}\right)^{-\widehat{\gamma}(t_1,t_2)} - U\left(\frac{1}{p}\right)\\
&= \left[X_k(t_1,t_2,1) - U\left(\frac{n}{k}\right)\right]\left(\frac{np}{k}\right)^{-\widehat{\gamma}(t_1,t_2)}\\
&\quad + \left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(t_1,t_2)} - \left(\frac{np}{k}\right)^{-\gamma}\right] U\left(\frac{n}{k}\right)\\
&\quad + \left[U\left(\frac{n}{k}\right)\left(\frac{np}{k}\right)^{-\gamma} - U\left(\frac{1}{p}\right)\right]\\
&=: I + II + III.
\end{aligned}$$

Before considering these three terms separately, observe that by the mean value theorem, using $\frac{\partial}{\partial\tau}(x^\tau) = x^\tau \log(x)$, there exists a $\nu \in [-1,1]$ such that

$$\begin{aligned}
\left(\frac{np}{k}\right)^{-\widehat{\gamma}(t_1,t_2)} - \left(\frac{np}{k}\right)^{-\gamma} &= (-\widehat{\gamma}(t_1,t_2)+\gamma)\left(\frac{np}{k}\right)^{-\gamma+\nu\left(\gamma-\widehat{\gamma}(t_1,t_2)\right)}\log\left(\frac{np}{k}\right)\\
&= \left(\frac{np}{k}\right)^{-\gamma}(\gamma-\widehat{\gamma}(t_1,t_2))\left(\frac{np}{k}\right)^{\nu\left(\gamma-\widehat{\gamma}(t_1,t_2)\right)}\log\left(\frac{np}{k}\right).
\end{aligned}$$

Combine this with

$$\begin{aligned}
\left(\frac{np}{k}\right)^{\nu\left(\gamma-\widehat{\gamma}(t_1,t_2)\right)} &= \exp\left[\nu\left(\gamma-\widehat{\gamma}(t_1,t_2)\right)\log\left(\frac{np}{k}\right)\right]\\
&\overset{(3.14)}{=} \exp\left[\nu\mathcal{O}_{\mathrm{P}}\left(\frac{1}{\sqrt{k}}\right)\log\left(\frac{np}{k}\right)\right] \overset{\textbf{(C5)}}{=} 1 + o_{\mathrm{P}}(1)
\end{aligned}$$

and **(B6)** to get

$$\begin{aligned}
&\frac{U\left(\frac{n}{k}\right)}{U\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(t_1,t_2)}-\left(\frac{np}{k}\right)^{-\gamma}\right]\\
=&\frac{U\left(\frac{n}{k}\right)}{U\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\gamma}\left(\gamma-\widehat{\gamma}(t_1,t_2)\right)\left(\frac{np}{k}\right)^{\nu\left(\gamma-\widehat{\gamma}(t_1,t_2)\right)}\log\left(\frac{np}{k}\right)\right]\\
=&\left(1+o_{\mathrm{P}}\left(1\right)\right)\left(\gamma-\widehat{\gamma}(t_1,t_2)\right)\log\left(\frac{np}{k}\right).
\end{aligned}\tag{3.17}$$

For the first term we obtain, using (3.13) in conjunction with the delta method (see also the proof of Theorem 3 in Einmahl *et al.*, 2016) for the fourth equality,

$$\begin{aligned}
&\frac{I}{U(1/p)\log\left(\frac{k}{np}\right)}\\
=\;&\frac{1}{U\left(\frac{1}{p}\right)}\left(\frac{np}{k}\right)^{-\widehat{\gamma}(t_1,t_2)}\frac{1}{\log\left(\frac{k}{np}\right)}\left[X_k(t_1,t_2,1)-U\left(\frac{n}{k}\right)\right]\\
=\;&\frac{U\left(\frac{n}{k}\right)}{U\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(t_1,t_2)}-\left(\frac{np}{k}\right)^{-\gamma}+\left(\frac{np}{k}\right)^{-\gamma}\right]\frac{1}{\log\left(\frac{k}{np}\right)}\left[\frac{X_k(t_1,t_2,1)}{U\left(\frac{n}{k}\right)}-1\right]\\
\underset{(3.17)}{\overset{\textbf{(B6)}}{=}}\;&\left[(1+o_{\mathrm{P}}(1))(\gamma-\widehat{\gamma}(t_1,t_2))\log\left(\frac{np}{k}\right)+1+o\left(\frac{1}{\sqrt{k}}\right)\right]\frac{\left[\frac{X_k(t_1,t_2,1)}{U\left(\frac{n}{k}\right)}-1\right]}{\log\left(\frac{k}{np}\right)}\\
=\;&\left[(1+o_{\mathrm{P}}(1))\sqrt{k}(\gamma-\widehat{\gamma}(t_1,t_2))k^{-1/2}\log\left(\frac{np}{k}\right)+1+o\left(\frac{1}{\sqrt{k}}\right)\right]\frac{\mathcal{O}_{\mathrm{P}}(\frac{1}{\sqrt{k}})}{\log\left(\frac{k}{np}\right)}\\
\underset{(3.15)}{\overset{\textbf{(B5)}}{=}}\;&o_{\mathrm{P}}\left(1/\sqrt{k}\right).
\end{aligned}\tag{3.18}$$

Further, utilizing **(B5)** and **(B6)** for the third term gives

$$\frac{III}{U(1/p)\log\left(\frac{k}{np}\right)}=\frac{1}{\log\left(\frac{k}{np}\right)}\left[\frac{U\left(\frac{n}{k}\right)}{U\left(\frac{1}{p}\right)\left(\frac{np}{k}\right)^{-\gamma}}-1\right]=o\left(1/\sqrt{k}\right).\tag{3.19}$$

The second term is non-negligible, since

$$\begin{aligned}\frac{II}{U(1/p)\log\left(\frac{k}{np}\right)} &= \frac{U\left(\frac{n}{k}\right)}{U\left(\frac{1}{p}\right)}\left(\left(\frac{np}{k}\right)^{-\widehat{\gamma}(t_1,t_2)}-\left(\frac{np}{k}\right)^{-\gamma}\right)\frac{1}{\log\left(\frac{k}{np}\right)}\\ &\overset{(3.17)}{=} \left(\widehat{\gamma}(t_1,t_2)-\gamma\right)\left(1+o_{\mathrm{P}}\left(1\right)\right). \end{aligned}\tag{3.20}$$

Combining (3.18)-(3.20) completes the proof. □

Proof of Theorem 3.2: We evaluate the numerator and denominator of Q_1^{EQ} (both premultiplied by the normalizing factor $\sqrt{k}/\log(k/(np))$ from Theorem 3.1) at t^*. Then the numerator converges to ∞ in probability because, using Theorem 3.1,

$$\begin{aligned}\frac{\sqrt{k}}{\log\left(\frac{k}{np}\right)}\log\left(\frac{\widehat{x}_p\left(0,t^*\right)}{\widehat{x}_p\left(t^*,1\right)}\right) &= \underbrace{\frac{\sqrt{k}}{\log\left(\frac{k}{np}\right)}\log\left(\frac{\widehat{x}_p\left(0,t^*\right)}{U_{\mathrm{pre}}(1/p)}\right)}_{=\mathcal{O}_P(1)}\\ &\quad-\underbrace{\frac{\sqrt{k}}{\log\left(\frac{k}{np}\right)}\log\left(\frac{\widehat{x}_p\left(t^*,1\right)}{U_{\mathrm{post}}(1/p)}\right)}_{=\mathcal{O}_P(1)}-\underbrace{\frac{\sqrt{k}}{\log\left(\frac{k}{np}\right)}\log\left(\frac{U_{\mathrm{post}}(1/p)}{U_{\mathrm{pre}}(1/p)}\right)}_{\to\pm\infty}.\end{aligned}$$

For the denominator we have, using Theorem 3.1 again,

$$\int_{t_0}^{t^*}\left[\frac{\sqrt{k}}{\log\left(\frac{k}{np}\right)}s\log\left(\frac{\widehat{x}_p\left(0,s\right)}{\widehat{x}_p\left(0,t^*\right)}\right)\right]^2\mathrm{d}s\ \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}}\ \sigma^2_{\widehat{\gamma},\gamma_{\mathrm{pre}}}\int_{t_0}^{t^*}\left[W(s)-\frac{s}{t^*}W(t^*)\right]^2\mathrm{d}s$$

and similarly for the other term. Hence, both terms are $\mathcal{O}_{\mathrm{P}}(1)$, whence the denominator is $\mathcal{O}_{\mathrm{P}}(1)$. The result follows. □

4 Sequential monitoring of the tail behavior of dependent data

We construct a sequential monitoring procedure for changes in the tail index and extreme quantiles of β-mixing random variables, which can be based on a large class of tail index estimators[1]. The assumptions on the data are general enough to be satisfied in a wide range of applications. In a simulation study empirical sizes and power of the proposed tests are studied for linear and non-linear time series. Finally, we use our results to monitor Bank of America stock log-losses from 2007 to 2012 and detect changes in extreme quantiles without an accompanying detection of a tail index break.

4.1 Motivation

The tail index of a random variable is arguably one of the most important parameters of its distribution: It determines some fundamental properties like the existence of moments, tail asymptotics of the distribution and the asymptotic behavior of sums and maxima. As a measure of tail thickness, the tail index is used in fields where heavy tails are frequently encountered, such as (re)insurance, finance, and teletraffic engineering (cf. Resnick, 2007, Sec. 1.3, and the references cited therein). Particularly in finance, the closely related extreme quantiles play a prominent role as a risk measure called Value-at-Risk (VaR).

The use of the variance as a risk measure has a long tradition in finance. Under Gaussianity the variance completely determines the tails of the distribution, which is no longer the case with heavy-tailed data. Hence, in order to assess the tail behavior of a time series, practitioners often estimate the tail index or an extreme quantile, the implicit assumption being their constancy over time. There are several suggestions in the literature on how to test this crucial assumption: Quintos *et al.* (2001) developed so called recursive, rolling and sequential tests for independent and GARCH data for tail index constancy based on the Hill (1975) estimator. Kim and Lee (2011) investigated their tests for more general β-mixing time series. Taking a likelihood approach for independent data, Dierckx and Teugels (2010) focus on breaks in the tail index for environmental data. Tests based on other estimators than the Hill

[1] This chapter was co-authored by Dominik Wied.

(1975) estimator were first proposed by Einmahl *et al.* (2016) for independent and in Chapter 2 for dependent data. To the best of our knowledge the only work dealing with changes in extreme quantiles is presented in Chapter 3. All these tests are of a retrospective nature.

We are not aware of any work on online surveillance methods for constancy of the tail index and extreme quantiles. This is important because, as noted in Chu *et al.* (1996), '[b]reaks can occur at any point, and given the costs of failing to detect them, it is desirable to detect them as rapidly as possible. One-shot tests cannot be applied in the usual way each time new data arrive, because repeated application of such tests yields a procedure that rejects a true null hypothesis of no change with probability one as the number of applications grows.' This chapter will fill this gap for closed-end procedures. To allow for sufficient flexibility in the use of tail index estimators, we will use the approach of Chapter 2.

Whether a monitoring procedure for a change in the tail index or an extreme quantile is of interest will largely be a matter of context. If interest centers on VaR, which is widely used in the banking industry and by financial regulators as a risk measure, the quantile monitoring procedure will be more relevant. If however interest centers on the mean excess function of the (log-transformed) data X, then, since $\mathrm{E}\left(\log X - \log t | X > t\right)$ converges to the extreme value index of X as $t \to \infty$, the tail index alternative seems more appropriate. Furthermore, the tail index per se could also be of interest as there are indications that it has predictive power for stock returns (Kelly and Jiang, 2014), where higher (lower) tail indices of returns indicate higher (lower) absolute returns.

The outline of this chapter is as follows. The main results under the null and two alternatives are stated in Section 4.2, where an example of a time series satisfying our assumptions is also given. Simulations and an empirical application are presented in Sections 4.3 and 4.4 respectively. All proofs are collected in Section 4.5.

4.2 Main results

4.2.1 Preliminaries and assumptions

To introduce the required notation let $X_1, \dots, X_n$ be a sequence of random variables defined on some probability space $(\Omega, \mathcal{A}, P)$ with survivor function $\bar{F}_i(x) := 1 - F_i(x) = \mathrm{P}\left(X_i > x\right)$, that is regularly varying with parameter $-\alpha_i$ (written $\bar{F}_i \in RV_{-\alpha_i}$), i.e.,

$$\bar{F}_i(x) = x^{-\alpha_i} L_i(x), \quad x > 0, \tag{4.1}$$

where $L_i : (0, \infty) \to (0, \infty)$ is slowly varying, i.e.,

$$\lim_{x \to \infty} \frac{L_i(\lambda x)}{L_i(x)} = 1 \quad \forall\, \lambda > 0. \tag{4.2}$$

If X_i is Pareto distributed, then $L_i(x) \equiv c > 0$. Since slow variation of the function $L_i(x)$ means, loosely speaking, that it behaves like a constant function at infinity, we say that X_i with tails as in (4.1) has *Pareto-type tails*. In the context of extreme value theory, α_i is called the tail index and $\gamma_i := 1/\alpha_i$ the extreme value index.

Define

$$U_i(x) := F_i^{-1}\left(1 - \frac{1}{x}\right), \quad x > 1,$$

as the $(1 - 1/x)$-quantile, F_i^{-1} being the left-continuous inverse of F_i. Then recall from Chapter 1 that (4.1) is equivalent to

$$\frac{U_i(\lambda x)}{U_i(x)} \underset{(x\to\infty)}{\longrightarrow} \lambda^\gamma. \tag{4.3}$$

Throughout, $k = k_n \in \mathbb{N}$ will denote a sequence satisfying $k \leq n - 1$,

$$k \underset{(n\to\infty)}{\longrightarrow} \infty \quad \text{and} \quad \frac{k}{n} \underset{(n\to\infty)}{\longrightarrow} 0, \tag{4.4}$$

controlling the number of upper order statistics used in the estimation of the tail index and $p = p_n \to 0$, $n \to \infty$, will denote a sequence of small probabilities, for which we want to test for a change in an appertaining extreme (right-tail) quantile $U_i(1/p)$. As is customary in extreme value theory, we will usually drop the subindex n and simply write k and p. For $t - s \geq 1/n$ and $y \in [0, 1]$ set

$$X_k(s, t, y) := \Big(\lfloor k(t-s)y\rfloor + 1\Big)\text{-th largest value of } X_{\lfloor ns\rfloor+1}, \ldots, X_{\lfloor nt\rfloor}.$$

Under the assumption of strictly stationary X_i we write $\bar{F} = \bar{F}_i$ and $U = U_i$. Let

$$\widehat{\gamma}(s,t) := \widehat{\gamma}_n(s,t), \qquad 0 \leq s < t < \infty,\ t - s \geq 1/n,$$

denote a generic tail index estimator based on the $\Big(\lfloor k(t-s)\rfloor + 1\Big)$-largest order statistics of the subsample $X_{\lfloor ns\rfloor+1}, \ldots, X_{\lfloor nt\rfloor}$. Then approximate with (4.3) for $x = 1/p$, $\lambda = pn/k$ with small $p > 0$ to get

$$\begin{aligned} U\left(\frac{1}{p}\right) &\approx U\left(\frac{n}{k}\right)\left(\frac{pn}{k}\right)^{-\gamma} \\ &\approx X_k(s,t,1)\left(\frac{np}{k}\right)^{-\widehat{\gamma}(s,t)} =: \widehat{x}_p(s,t), \end{aligned} \tag{4.5}$$

which motivates and defines the so called Weissman (1978) estimator for the extreme quantile $U(1/p)$. Hence, the idea is to use the (within sample range) estimator $X_k(s,t,1)$ of $U(n/k)$ to estimate the (possibly out of sample range) quantile $U(1/p)$

by exploiting the regular variation of the quantile function in (4.3). In view of this we will require $p \ll k/n$.

For concreteness we will focus in the following on the Hill (1975) estimator of γ given by

$$\widehat{\gamma}_H(0,1) := \frac{1}{k}\sum_{i=0}^{k-1} \log\left(\frac{X_{n-i:n}}{X_{n-k:n}}\right), \tag{4.6}$$

where $X_{n:n} \geq X_{n-1:n} \geq \ldots \geq X_{1:n}$ denote the order statistics of the sample $X_1, \ldots, X_n$. But the main results in Theorems 4.1 and 4.2 below hold to the letter for the moments-ratio estimator of Daníelsson *et al.* (1996) and the class of estimators introduced by Csörgő and Viharos (1998), see also the proof of Theorem 4.1 below.

The dependence concept we will use in the following is that of β-mixing, that is for a sequence of random variables $\{X_i\}_{i\in\mathbb{N}}$ the β-mixing coefficients $\beta(l)$ converge to zero:

$$\beta(l) := \sup_{m\in\mathbb{N}} \mathbb{E}\left[\sup_{A\in\mathcal{F}_{m+l+1}^{\infty}} \left|\mathrm{P}\left(A|\mathcal{F}_1^m\right) - \mathrm{P}\left(A\right)\right|\right] \underset{(l\to\infty)}{\longrightarrow} 0,$$

where $\mathcal{F}_m^{\infty} := \sigma\left(X_m, X_{m+1}, \ldots\right)$ and $\mathcal{F}_l^m := \sigma\left(X_l, \ldots, X_m\right)$; see also Chapter 1.

Write $D[a,b]$ for the space of cadlag functions on $[a,b]$ $(0 \leq a < b < \infty)$ endowed with the Skorohod topology and $D(I)$, $I \subset \mathbb{R}^m$ compact, for the multiparameter extension.

In order to construct a monitoring procedure for a tail index change, we have to assume tail index (or, equivalently, extreme value index) constancy over some historical period (also called training period) of suitable length n:

(C1) $\gamma_1 = \ldots = \gamma_n, \qquad n \in \mathbb{N}.$

This assumption can of course be tested by any of the retrospective change point tests proposed in Chapters 2 and 3.

As soon as a period $X_1, \ldots, X_n$ of tail index or extreme quantile stability is identified and more observations $X_{n+1}, X_{n+2}, \ldots$ become available, we are interested in an online surveillance method testing

$$\begin{aligned} \mathcal{H}_{0,\gamma}: &\quad \gamma_1 = \ldots = \gamma_n = \gamma_{n+1} = \ldots \quad \text{vs.} \\ \mathcal{H}_{1,\gamma}^{\lessgtr}: &\quad \gamma_1 = \ldots = \gamma_n = \ldots = \gamma_{\lfloor nt^*\rfloor} \lessgtr \gamma_{\lfloor nt^*\rfloor+1} = \gamma_{\lfloor nt^*\rfloor+2} = \ldots, \end{aligned} \tag{4.7}$$

and

$$
\begin{aligned}
\mathcal{H}_{0,U}: \quad & U_1\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right) = U_{n+1}\left(\frac{1}{p}\right) = \ldots \quad \text{vs.} \\
\mathcal{H}_{1,U}^{\lessgtr}: \quad & U_1\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right) = \ldots = U_{\lfloor nt^* \rfloor}\left(\frac{1}{p}\right) \lessgtr \\
& \qquad\qquad U_{\lfloor nt^* \rfloor+1}\left(\frac{1}{p}\right) = U_{\lfloor nt^* \rfloor+2}\left(\frac{1}{p}\right) = \ldots
\end{aligned} \tag{4.8}
$$

for some $t^* \geq 1$ denoting the unknown change point. We use $\mathcal{H}_0$ or $\mathcal{H}_1^{\lessgtr}$ as shorthand notation for both of $\mathcal{H}_{0,\gamma}$, $\mathcal{H}_{0,U}$ or $\mathcal{H}_{1,\gamma}^{\lessgtr}$, $\mathcal{H}_{1,U}^{\lessgtr}$.

We use the following detectors for (4.7)

$$
\begin{aligned}
V_n^{\widehat{\gamma}}(t) &:= \frac{\left[(t-1)\left(\widehat{\gamma}(1,t) - \widehat{\gamma}(0,1)\right)\right]^2}{\int_{t_0}^1 \left[s\left(\widehat{\gamma}(0,s) - \widehat{\gamma}(0,1)\right)\right]^2 \mathrm{d}s}, & t \geq 1 + t_0, \\
W_n^{\widehat{\gamma}}(t) &:= \frac{\left[t_0\left(\widehat{\gamma}(t-t_0,t) - \widehat{\gamma}(0,1)\right)\right]^2}{\int_{t_0}^1 \left[t_0\left(\widehat{\gamma}(s-t_0,s) - \widehat{\gamma}(0,1)\right)\right]^2 \mathrm{d}s}, & t \geq 1 + t_0,
\end{aligned}
$$

and for (4.8)

$$
\begin{aligned}
V_n^{\widehat{x}_p}(t) &:= \frac{\left[(t-1)\log\left(\frac{\widehat{x}_p(1,t)}{\widehat{x}_p(0,1)}\right)\right]^2}{\int_{t_0}^1 \left[s\log\left(\frac{\widehat{x}_p(0,s)}{\widehat{x}_p(0,1)}\right)\right]^2 \mathrm{d}s}, & t \geq 1 + t_0, \\
W_n^{\widehat{x}_p}(t) &:= \frac{\left[t_0\log\left(\frac{\widehat{x}_p(t-t_0,t)}{\widehat{x}_p(0,1)}\right)\right]^2}{\int_{t_0}^1 \left[t_0\log\left(\frac{\widehat{x}_p(s-t_0,s)}{\widehat{x}_p(0,1)}\right)\right]^2 \mathrm{d}s}, & t \geq 1 + t_0,
\end{aligned}
$$

where $t_0 > 0$ defines the (minimal) fraction of n upon which the tail index and extreme quantile estimators are based. To motivate our detectors consider $V_n^{\widehat{\gamma}}$, the others can be motivated similarly. In the numerator the training period estimate $\widehat{\gamma}(0,1)$ is compared with the current observation period estimate $\widehat{\gamma}(1,t)$. If the observation period length $(t-1)$ is large, that difference is weighted more heavily. The denominator 'self-normalizes' the numerator. While we could have chosen a wide range of functionals for this (e.g., the denominator of $W_n^{\widehat{\gamma}}$), it seemed more natural to incorporate the functional form of the numerator to do so. With this motivation in mind we are inclined to reject $\mathcal{H}_0$ if the following stopping times terminate (in the sense of being finite):

$$\tau_n^{V^{\widehat{\gamma}}} := \inf\left\{t \in [1+t_0, T] \; : \; V_n^{\widehat{\gamma}}(t) > c\right\},$$
$$\tau_n^{W^{\widehat{\gamma}}} := \inf\left\{t \in [1+t_0, T] \; : \; W_n^{\widehat{\gamma}}(t) > c\right\},$$

and

$$\tau_n^{V^{\widehat{x}_p}} := \inf\left\{t \in [1+t_0, T] \; : \; V_n^{\widehat{x}_p}(t) > c\right\},$$
$$\tau_n^{W^{\widehat{x}_p}} := \inf\left\{t \in [1+t_0, T] \; : \; W_n^{\widehat{x}_p}(t) > c\right\},$$

where from now on $\inf \emptyset := \infty$ and $c > 0$ is chosen, such that under $\mathcal{H}_0$,

$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n < \infty\right) = \alpha$$

for some prespecified significance level $\alpha \in (0,1)$ (see Theorem 4.1 below). Here $T > 1$ denotes the arbitrarily large closed end of the procedure, i.e., the method terminates after observations $X_{n+1}, \ldots, X_{\lfloor nT \rfloor}$. Closed-end procedures are quite common, e.g., Aue *et al.* (2012) consider breaks in portfolio betas, Wied and Galeano (2013) breaks in cross-correlations, Zeileis *et al.* (2005) and Aschersleben *et al.* (2015) breaks in regression and cointegrating relationships respectively.

Remark 4.1. (a) The detector $V_n^{\widehat{\gamma}}$ comes closer in spirit to many of the detectors in the online monitoring literature, where an estimate of some parameter based on the historical period is compared to that based on the (ever longer) monitoring period, see the references in the paragraph before. However, the procedure based on $V_n^{\widehat{\gamma}}$ is not consistent against $\mathcal{H}_{1,\gamma}^{>}$, cf. Theorem 3.2 below, which is the reason for introducing the method based on $W_n^{\widehat{\gamma}}$.

(b) We could have based our procedure equally well on detectors of the type

$$V_n^{\widehat{\gamma}}(t) := \frac{1}{\sigma^2_{\widehat{\gamma},\gamma}} \left[(t-1)\sqrt{k}\left(\widehat{\gamma}\left(1,t\right) - \widehat{\gamma}\left(0,1\right)\right)\right]^2,$$

where $\sigma^2_{\widehat{\gamma},\gamma}$ is a consistent estimator of the asymptotic variance of $\sqrt{k}(\widehat{\gamma}(0,1)-\gamma)$ based on the observations $X_1, \ldots, X_n$ in the observation period (e.g., the one in Theorem 2.2). It turns out however, that in simulations values of even $n = 500$ for the training period were not sufficient to deliver accurate variance estimates for a wide range of model parameters for the models we investigated, which lead to severe size distortions of our surveillance methods. This is why we opted for the self-normalized approach advocated in Shao and Zhang (2010) in our sequential setting. To the best of our knowledge we are the first to do so. Shao and Zhang (2010) found that for retrospective change point tests self-normalized test statistics delivers far superior size in simulations. However, the price to be paid for using a self-normalization approach versus a variance estimation approach is slightly lower power.

Under $\mathcal{H}_0$ we will assume beyond **(C1)** that:

(C2) $\{X_i\}_{i\in\mathbb{N}}$ is a strictly stationary β-mixing process with continuous marginals and mixing coefficients $\beta(\cdot)$, such that

$$\lim_{n\to\infty} \frac{n}{r_n}\beta(l_n) + \frac{r_n}{\sqrt{k}}\log^2(k) = 0$$

for sequences $\{l_n\}_{n\in\mathbb{N}} \subset \mathbb{N}, \{r_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$ tending to infinity with $l_n = o(r_n)$, $r_n = o(n)$.

(C3) There exists a function $r = r(x, y)$, such that for all $x, y \in [0, y_0 + \varepsilon]$ $(y_0 \geq 1, \varepsilon > 0)$

$$\lim_{n\to\infty} \frac{n}{r_n k} \sum_{1\leq i,j\leq r_n} \operatorname{Cov}\left(I_{\left\{X_i > U\left(\frac{n}{kx}\right)\right\}}, I_{\left\{X_j > U\left(\frac{n}{ky}\right)\right\}}\right) = r(x, y).$$

(C4) For some constant $C > 0$

$$\frac{n}{r_n k}\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U\left(\frac{n}{ky}\right) < X_i \leq U\left(\frac{n}{kx}\right)\right\}}\right]^4 \leq C(y - x) \quad \forall\ 0 \leq x < y \leq y_0 + \varepsilon,\ n \in \mathbb{N}.$$

(C5) There exist $\rho < 0$ and a function $A(\cdot)$ with eventually positive or negative sign and $\lim_{t\to\infty} A(t) = 0$, such that for any $y > 0$

$$\lim_{t\to\infty} \frac{\frac{U(ty)}{U(t)} - y^\gamma}{A(t)} = y^\gamma \frac{y^\rho - 1}{\rho},$$

where $\sqrt{k}A(n/k) \to 0$ as $n \to \infty$.

For the detectors for changes in extreme quantiles we need the following further assumptions:

(C6) $\lim_{n\to\infty} \frac{np}{k} = 0$, $\lim_{n\to\infty} k^{-1/2} \log(np) = 0$.

(C7) The sequence k satisfies

$$\frac{U(1/p)}{U(n/k)} \left(\frac{np}{k}\right)^{\gamma} - 1 = o\left(\frac{1}{\sqrt{k}}\right).$$

Conditions **(C2)**-**(C7)** are quite similar to **(B1)**-**(B6)** from Chapter 3. The conditions **(C2)**-**(C4)** correspond (almost) exactly to conditions ($\widetilde{\mathrm{C1}}$), ($\widetilde{\mathrm{C2}}$) and ($\widetilde{\mathrm{C3}^*}$) in Drees (2000). Condition **(C5)**, which is a widely used second-order strengthening of (4.3) (e.g., Kim and Lee, 2011; Einmahl *et al.*, 2016), is stronger than Drees's (2000) condition (3.5). **(C2)** ensures a standard 'big block - small block' approach may be applied to deduce weak convergence of what Einmahl *et al.* (2016, p. 42) termed the *simple sequential tail empirical process* in (4.18) below. The limit process has a well-defined covariance structure by virtue of **(C3)**. **(C6)** provides a range for p: $\lim_{n\to\infty} \frac{np}{k} = 0$ provides an upper bound for the decay of p (indicating the limitations of the extreme value theory approach towards the center of the distribution) while $\lim_{n\to\infty} k^{-1/2} \log(np/k) = \lim_{n\to\infty} k^{-1/2} \log(np) = 0$ provides a lower bound, beyond which estimation is no longer feasible. Note that if the d.f. obeys the quite general expansion

$$1 - F(x) = Cx^{-\alpha}(1 + \mathcal{O}(x^{-\beta})) \quad \text{as} \quad x \to \infty; \quad C, \alpha, \beta > 0,$$

then by inversion $U(x) = C^{1/\alpha} x^{1/\alpha}(1 + \mathcal{O}(x^{-\beta/\alpha}))$, whence **(C7)** does not impose an additional constraint on the choice of k.

4.2.2 Results under the null and the alternative

We are now ready to state the main results under the null, describing the asymptotic behavior of our monitoring procedures based on the stopping rules τ_n.

Theorem 4.1. *Suppose* **(C1)**-**(C5)** *hold for* $y_0 = 1$. *Then for any* $t_0 > 0$, $T > 1 + t_0$ *and*

$$\begin{aligned} V_{t_0,T} &:= \frac{\sup_{t\in[1+t_0,T]} \left[W(t) - tW(1)\right]^2}{\int_{t_0}^1 \left[W(s) - sW(1)\right]^2 \mathrm{d}s}, \\ W_{t_0,T} &:= \frac{\sup_{t\in[1+t_0,T]} \left[W(t) - W(t-t_0) - t_0 W(1)\right]^2}{\int_{t_0}^1 \left[W(s) - W(s-t_0) - t_0 W(1)\right]^2 \mathrm{d}s}, \end{aligned} \tag{4.9}$$

with $\{W(t)\}_{t\in[0,T]}$ *a standard Brownian motion,* *(i) under $\mathcal{H}_{0,\gamma}$*

$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{\widehat{V^\gamma}} < \infty\right) = \mathrm{P}\left(V_{t_0,T} > c\right),$$
$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{\widehat{W^\gamma}} < \infty\right) = \mathrm{P}\left(W_{t_0,T} > c\right),$$

(ii) under $\mathcal{H}_{0,U}$ and additionally **(C6)**-**(C7)**

$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{\widehat{V^{x_p}}} < \infty\right) = \mathrm{P}\left(V_{t_0,T} > c\right),$$
$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{\widehat{W^{x_p}}} < \infty\right) = \mathrm{P}\left(W_{t_0,T} > c\right).$$

Next, we investigate the behavior of our procedures under the 'one-sided' alternatives $\mathcal{H}_1^{\lessgtr}$. To prove our results the observations will be denoted by the triangular array of random variables $X_{n,i}$, $n \in \mathbb{N}$, $i = 1, \ldots, n$, which have a common marginal survivor function $\overline{F}_{\text{pre}} \in RV_{-\alpha_{\text{pre}}}$ ($\overline{F}_{\text{post}} \in RV_{-\alpha_{\text{post}}}$) for $i \in I_{\text{pre}} := \{1, \ldots, \lfloor nt^* \rfloor\}$ ($i \in I_{\text{post}} := \{\lfloor nt^* \rfloor + 1, \ldots, \lfloor nT \rfloor\}$). Set

$$U_{\text{pre}}(x) = F_{\text{pre}}^{-1}\left(1 - \frac{1}{x}\right) \quad \text{and} \quad U_{\text{post}}(x) = F_{\text{post}}^{-1}\left(1 - \frac{1}{x}\right).$$

Theorem 4.2. *Let the triangular array $\{X_{n,i}\}_{n\in\mathbb{N}, i=1,\ldots,n}$ be given by*

$$X_{n,i} := \begin{cases} Y_n, & i \in I_{\text{pre}}, \\ Z_n, & i \in I_{\text{post}}, \end{cases}$$

where $\{Y_n\}_{n\in\mathbb{N}}$ and $\{Z_n\}_{n\in\mathbb{N}}$ both satisfy conditions **(C2)**-**(C5)** *with $y_0 = T$ and*

$$k, \gamma_{\text{pre}}, U_{\text{pre}}(\cdot), r_{\text{pre}}(\cdot,\cdot) \qquad \textit{and} \qquad k, \gamma_{\text{post}}, U_{\text{post}}(\cdot), r_{\text{post}}(\cdot,\cdot)$$

respectively. Then

(i)

$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{\widehat{V^\gamma}} < \infty\right) = 1 \qquad \textit{under } \mathcal{H}_{1,\gamma}^{<},$$
$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{\widehat{V^\gamma}} < \infty\right) < 1 \qquad \textit{under } \mathcal{H}_{1,\gamma}^{>},$$

(ii)

$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{W\widehat{\gamma}} < \infty\right) = 1 \qquad \textit{under } \mathcal{H}_{1,\gamma}^{\lessgtr},$$

where under $\mathcal{H}_1^{>}$ additionally $t^* \in [1, T - t_0]$ must hold.

Suppose that additionally **(C6)**-**(C7)** *hold for $\{Y_n\}$ and $\{Z_n\}$. Then*

(iii)

$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{V\widehat{x_p}} < \infty\right) = 1 \qquad \textit{under } \mathcal{H}_{1,\gamma}^{\lessgtr},$$

(iv) if $t^* \in [1, T - t_0]$ and $\sqrt{k}/\log(k/(np))\log(U_{\mathrm{pre}}(1/p)/U_{\mathrm{post}}(1/p)) \underset{(n\to\infty)}{\longrightarrow} \infty$,

$$\lim_{n\to\infty} \mathrm{P}\left(\tau_n^{W\widehat{x_p}} < \infty\right) = 1 \qquad \textit{under } \mathcal{H}_{1,U}^{\lessgtr}.$$

Remark 4.2. (a) Note that the sequence k in the pre- and post-break period must be the same. This is however not too restrictive as (see (4.11) below) one can frequently choose $k = n^\nu$ for some arbitrarily small $\nu > 0$. So if **(C2)**-**(C7)** are satisfied for $k = n^\nu$ with $\nu \in (0, \nu_{\mathrm{pre}})$ for the pre-break period and with $\nu \in (0, \nu_{\mathrm{post}})$ in the post-break period, then taking $\nu \in (0, \min(\nu_{\mathrm{pre}}, \nu_{\mathrm{post}}))$ leads to a sequence for which all assumptions are satisfied for the whole sample.

(b) If the $X_{n,i}$ are generated as in Theorem 4.2, the hypothesis $\mathcal{H}_{1,\gamma}^{\lessgtr}$ is a strict subset of the hypothesis $\mathcal{H}_{1,U}^{\lessgtr}$. E.g., taking $Z_n = aY_n$, $a \neq 1$, is covered under $\mathcal{H}_{1,U}^{\lessgtr}$, but not under $\mathcal{H}_{1,\gamma}^{\lessgtr}$, since scaling does not affect the tail index. If however (e.g.) $\mathcal{H}_{1,\gamma}^{>}$ is true, we have $U_{\mathrm{pre}}/U_{\mathrm{post}} \in RV_{\gamma_{\mathrm{pre}}-\gamma_{\mathrm{post}}>0}$ and hence

$$\frac{U_{\mathrm{pre}}(1/p_n)}{U_{\mathrm{post}}(1/p_n)} \underset{(n\to\infty)}{\longrightarrow} \infty,$$

s.t. $\mathcal{H}_{1,U}^{>}$ is true.

(c) Under $\mathcal{H}_1^{>}$ the procedure based on $\tau_n^{V\widehat{\gamma}}$ is not consistent, which motivated the study of $\tau_n^{W\widehat{\gamma}}$. The reason for the inconsistency is, simply speaking, that in a sample with one extreme value index break, extreme value index estimators will consistently estimate the larger extreme value index. Hence, if there is a break in t^* toward lighter tails in the observation period, then $\widehat{\gamma}(1, t)$, $t > t^*$, will still estimate the larger extreme value index, even though the last part of the sample upon which it is based possess a smaller extreme value index. Thus, the change goes unnoticed.

4.2.3 An example

In this subsection we verify the conditions **(C1)-(C4)** for the following stochastic volatility model

$$X_i = \sigma_i Z_i, \qquad i \in \mathbb{Z},$$

where $\{Z_i\}$ are i.i.d. and independent from $\{\sigma_i\}$. Denote by $1 - F_{|Z|} \in RV_{-\alpha}$, $\alpha > 0$, the survivor function of $|Z_0|$. The volatility process is assumed to be generated according to

$$\sigma_i = \exp(Y_i), \qquad \text{where } Y_i = \sum_{j=0}^{\infty} \psi_j \epsilon_{i-j},$$

with (w.l.o.g.) $\Psi_0 = 1$, $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ and geometrically decreasing coefficients $|\psi_j| \underset{(j\to\infty)}{=} \mathcal{O}(\eta^j)$, $\eta \in (0,1)$, covering all finite order ARMA(p,q)-models. Many popular stochastic volatility models use (zero-mean) Gaussian AR(1)-models for the volatility dynamics (Asai *et al.*, 2006).

We now verify our conditions for $|X_i| = \sigma_i |Z_i|$. For **(C2)** strict stationarity is immediate (which also implies **(C1)**). By Bradley (1986, Ex. 6.1) the Y_i are geometrically β-mixing, whence, by Gaussianity of the Y_i, they are also geometrically ρ-mixing (Bradley, 1986, Eq. (1.7) & Thm. 5.1 and the comments below it), i.e., the mixing coefficients

$$\rho(j) = \sup_{\substack{U \in \mathcal{L}_2(\mathcal{B}_{-\infty}^0) \\ V \in \mathcal{L}_2(\mathcal{B}_j^\infty)}} \left|\operatorname{corr}(U, V)\right| \underset{(j\to\infty)}{\longrightarrow} 0 \tag{4.10}$$

decay to zero geometrically fast. Here, $\mathcal{B}_{-\infty}^0 = \sigma(\ldots, Y_{-1}, Y_0)$, $\mathcal{B}_j^\infty = \sigma(Y_j, Y_{j+1}, \ldots)$ and $\mathcal{L}_2(\mathcal{A})$ is the space of square-integrable, $\mathcal{A}$-measurable real-valued functions. By Remark 1.15 this implies geometric ρ-mixing of $\sigma_i = \exp(Y_i)$. This in turn implies geometric ρ-mixing of $X_i = \sigma_i Z_i$ since for $U \in \mathcal{L}_2(\sigma(\ldots, X_{-1}, X_0))$ and $V \in \mathcal{L}_2(\sigma(X_j, X_{j+1}, \ldots))$

$$\begin{aligned}
\mathrm{E}[UV] &= \mathrm{E}\left[\mathrm{E}[U|\sigma_s, s \le 0]\,\mathrm{E}[V|\sigma_s, s \ge n]\right], \\
\mathrm{E}[U] &= \mathrm{E}\left[\mathrm{E}[U|\sigma_s, s \le 0]\right], \\
\mathrm{E}[U^2] &= \mathrm{E}\left[\mathrm{E}[U^2|\sigma_s, s \le 0]\right] \ge \mathrm{E}\left[\mathrm{E}[U|\sigma_s, s \le 0]\right]^2,
\end{aligned}$$

where the first line follows from the independence of $\{\sigma_i\}$ and $\{Z_i\}$, the second from the law of iterated expectations and the third from the Cauchy-Schwarz inequality.

Hence, by the definition in (4.10), the mixing coefficients of X_i are bounded by those of σ_i, such that geometric ρ-mixing is inherited from the volatility process. We conclude for the mixing coefficients $\rho(\cdot)$ appertaining to the $|X_i|$ that $\rho(j) \leq K\eta^j$ for some $\eta \in (0,1)$. By Remark 1.15 and recalling that the Y_i are geometrically β-mixing, the same holds true for the β-mixing coefficients. Thus, **(C2)** is satisfied for the following choices

$$k = n^{\nu} \text{ for } \nu \in (0, 3/4), \quad l_n = -2\frac{\log n}{\log \eta}, \quad r_n = n^{\nu/3}. \tag{4.11}$$

We check conditions (C2) and (C3) of Drees (2003, Prop. 2.1), which implies **(C3)** because with the above choices $r_n k/n = o(1)$. For (C2) we get from Hill (2011a, Thm. 2.1) that as $n \to \infty$

$$\begin{aligned}
&P(|X_1| > xU(n/k), |X_{1+m}| > yU(n/k)) \\
\sim&\frac{\mathrm{E}[\sigma_1^\alpha \sigma_{1+m}^\alpha]}{\mathrm{E}[\sigma_1^\alpha]\,\mathrm{E}[\sigma_{1+m}^\alpha]} P(|X_1| > xU(n/k))P(|X_{1+m}| > yU(n/k)) \\
\sim&\frac{\mathrm{E}[\sigma_1^\alpha \sigma_{1+m}^\alpha]}{\mathrm{E}[\sigma_1^\alpha]\,\mathrm{E}[\sigma_{1+m}^\alpha]} \frac{k}{n} x^{-\alpha} \frac{k}{n} y^{-\alpha},
\end{aligned}$$

where the last line follows from (4.12) below and (e.g.) Resnick (2007, Sec. 2.2.1). (C3) is satisfied due to geometric ρ-mixing of X_i (see also Drees, 2003, Rem. 2.2). Assumption **(C4)** is again a consequence of Drees (2003, Prop. 2.1 & Rem. 2.3).

The X_i inherit their heavy tails from the Z_i as, by Breiman's (1965) lemma,

$$1 - F(x) = P(|X_0| > x) \underset{(x\to\infty)}{\sim} \mathrm{E}[\sigma_1^\alpha] P(|Z_0| > x). \tag{4.12}$$

Hence, the $|X_i|$ also have tail index α and (4.1) is satisfied. Of course, (4.12) only gives $1 - F(x) = cx^{-\alpha}(1 + o(1))$. This is weaker than

$$\lim_{t\to\infty} \frac{\frac{1-F(ty)}{1-F(t)} - y^{-\alpha}}{A(1/[1-F(t)])} = y^{-\alpha}\frac{y^{\rho\alpha} - 1}{\rho/\alpha}, \tag{4.13}$$

with $\sqrt{k}A(n/k) \to 0$ as $n \to \infty$, which is equivalent to **(C5)** (cf. de Haan and Ferreira, 2006, Thm. 2.3.9). The currently sharpest result on the second-order behavior of the d.f.s of stochastic volatility models seems to be Kulik and Soulier (2011, Prop. 2.8). Assume

$$1 - F_{|Z|}(x) = c_z x^{-\alpha} \exp\left(\int_1^x \frac{\eta(s)}{s}\mathrm{d}s\right), \qquad x > 0,\ c > 0$$

for some $\eta(s) = \mathcal{O}(s^{\alpha\rho})$, $\rho < 0$. For instance Fréchet-, $|t_\nu|$- and generalized Pareto distributions have such tails (cf. Beirlant *et al.*, 2004, Sec. 2.3.4). Then the aforementioned proposition implies

$$1 - F(x) = cx^{-\alpha}(1 + \mathcal{O}(x^{\alpha\rho})), \qquad x > 0,\ c > 0, \tag{4.14}$$

which is stronger than $1 - F(x) = cx^{-\alpha}(1 + o(1))$ from (4.12), but not quite sufficient for (4.13). However, if the $\mathcal{O}$-term in (4.14) satisfied an expansion $c_1 x^{\alpha\rho}(1 + o(1))$, then (4.13) would be satisfied for $k = o(n^{2\rho/(2\rho-1)})$ (cf. de Haan and Ferreira, 2006, p. 77). Recall from the discussion of conditions **(C2)**-**(C7)** that (4.14) implies **(C7)**.

4.3 Simulations

In this section we investigate the finite-sample behavior of the monitoring procedures based on the stopping times $\tau_n^{W^{\widehat{\gamma}/\widehat{x}_p}}$ and $\tau_n^{V^{\widehat{\gamma}/\widehat{x}_p}}$. Throughout we simulate 10,000 time series with historical periods of length $n = 125, 500$ and $T = 4$, such that the total length is $\lfloor nT \rfloor = 500, 2000$. Furthermore we use $t_0 = 0.2$ and $k/n = 0.2$ and the estimator $\widehat{\gamma} = \widehat{\gamma}_H$ of the extreme value index we employ is the Hill (1975) estimator. Simulation results were quite robust to the particular choice of k/n and are available from the authors upon request. The quantiles of the distributions of $V_{t_0,T}$ and $W_{t_0,T}$ from (4.9) are tabulated in Table 4.1 for $t_0 = 0.2$ and $T = 4$. To simulate them we used 100,000 realizations of Brownian motions on the interval $[0, 4]$, which themselves were generated using 400,000 normally distributed random variables.

α_q	0.50	0.60	0.70	0.80	0.90	0.95	0.99
α_q-quantile of $V_{t_0,T}$	78.35	113.0	166.2	257.8	464.5	723.4	1557
α_q-quantile of $W_{t_0,T}$	15.57	18.42	22.10	27.39	36.79	46.87	72.89

Table 4.1: Quantiles of $V_{t_0,T}$ and $W_{t_0,T}$ ($t_0 = 0.2$, $T = 4$)

We investigate size using data from a linear ARMA(1,1) and non-linear ARCH(1) and SV models. Specifically, we simulate $\{X_i\}_{i=1,\ldots,\lfloor nT \rfloor}$ from the three data generating processes (DGPs)

$$X_i = 0.3 \cdot X_{i-1} + Z_i + 0.7 \cdot Z_{i-1}, \qquad Z_i \overset{\text{i.i.d.}}{\sim} t_{10}, \tag{ARMA}$$

$$X_i^2 = (0.01 + 0.3125 \cdot X_{i-1}^2) Z_i^2, \qquad Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{ARCH}$$

$$|X_i| = \sigma_i |Z_i|, \qquad Z_i \overset{\text{i.i.d.}}{\sim} t_{0.5}, \tag{SV}$$

where t_ν denotes a Student's t-distribution with ν degrees of freedom (i.e., tail index equal to ν) and $\sigma_i = 0.5\sigma_{i-1} + \varepsilon_i$, $\varepsilon \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, a Gaussian AR(1)-process. For the

X	DGP	Size	$\lfloor nT \rfloor = 500$				$\lfloor nT \rfloor = 2000$			
			$X^{\widehat{\gamma}}$	$X^{\widehat{x}_p}$			$X^{\widehat{\gamma}}$	$X^{\widehat{x}_p}$		
				0.1	0.01	0.001		0.1	0.01	0.001
V	(ARMA)	10	5.4	11	7.7	6.0	7.1	10	8.7	7.9
		5	2.7	6.2	4.0	3.0	3.6	5.5	4.5	3.7
	(ARCH)	10	5.9	12	9.0	6.9	7.5	11	9.0	8.0
		5	2.8	6.6	4.6	3.5	3.8	5.5	4.5	4.1
	(SV)	10	8.0	11	10	9.2	8.5	10	10	9.3
		5	4.3	5.9	6.0	5.3	4.3	5.5	5.1	4.6
W	(ARMA)	10	15	8.4	11	14	10	8.6	8.8	9.7
		5	8.8	4.4	6.7	8.3	5.5	4.4	4.7	4.9
	(ARCH)	10	13	7.9	12	13	11	8.5	11	12
		5	7.7	4.4	7.5	8.2	6.1	4.6	6.5	6.8
	(SV)	10	16	8.7	15	16	10	8.2	10	11
		5	11	5.2	10	11	5.8	4.0	5.7	5.9

Table 4.2: Empirical sizes in % of monitoring procedures based on $X^{\widehat{\gamma}}$ and $X^{\widehat{x}_p}$ ($X \in \{V, W\}$, $p \in \{0.1, 0.01, 0.001\}$) for $\lfloor nT \rfloor$ realizations of (ARMA), (ARCH) and (SV)

verification of the conditions **(C2)-(C7)** for the first two models we refer to Drees (2003, Secs. 3.1 & 3.2). The $|X_i|$ generated from the ARMA(1,1)-model have tail index 10 because of Lemma 5.2 in Datta and McCormick (1998). The tail index of the X_i^2 from the ARCH(1)-model is given by $8/2 = 4$ (cf. Davis and Mikosch, 1998, Table 1), while that of $|X_i|$ from (SV) is 0.5 (see Section 4.2.3). The parameters are chosen to demonstrate that our procedure works well for tails ranging from rather light (in the (ARMA) case) to very heavy with non-existent first moment (in the (SV) case).

The conclusions that can be drawn from Table 4.2 are quite similar for all models. When $\lfloor nT \rfloor = 500$ size varies around the nominal level quite a bit for different choices of p. This is no longer the case for the longer period, where size is always very close to the expected level. Interestingly, when the procedures based on the detectors V are oversized, those based on W are undersized and vice versa.

To assess the power of our tests we generate data from the following models, where the historical data in all three cases are generated according to the models already

X	DGP	t^*	Level	$\lfloor nT \rfloor = 500$				$\lfloor nT \rfloor = 2000$			
				$W^{\widehat{\gamma}}$	$W^{\widehat{x_p}}$			$W^{\widehat{\gamma}}$	$W^{\widehat{x_p}}$		
					0.1	0.01	0.001		0.1	0.01	0.001
V	(ARMA)	1.15	10	7.4	96	66	33	9.7	100	97	75
			5	3.4	92	53	23	4.5	100	94	62
		2.5	10	18	81	57	38	40	100	95	82
			5	11	70	45	27	27	98	89	71
	(ARCH)	1.15	10	17	43	37	28	37	77	71	59
			5	11	33	27	19	26	65	59	47
		2.5	10	12	26	23	18	22	44	42	34
			5	6.7	18	16	11	14	31	31	23
	(SV)	1.15	10	27	73	57	45	83	99	97	94
			5	15	58	40	29	69	97	91	85
		2.5	10	8.5	28	15	11	22	72	46	36
			5	3.6	17	7.4	5.0	11	57	32	22
W	(ARMA)	1.15	10	20	54	41	32	22	98	80	58
			5	13	41	31	23	14	95	68	45
		2.5	10	18	45	31	25	19	97	70	47
			5	11	33	22	17	12	92	57	34
	(ARCH)	1.15	10	32	30	43	40	50	60	69	64
			5	23	22	33	30	38	48	58	52
		2.5	10	25	22	31	30	37	47	54	49
			5	17	15	23	21	27	35	43	38
	(SV)	1.15	10	0.9	10	1.5	0.8	6.1	54	26	16
			5	0.4	5.3	0.6	0.0	2.2	38	14	7.8
		2.5	10	10	11	9.7	10	10	48	25	18
			5	6.2	5.6	5.9	6.1	4.9	33	14	9.2

Table 4.3: Empirical power in % of monitoring procedures based on $X^{\widehat{\gamma}}$ and $X^{\widehat{x_p}}$ ($X \in \{V, W\}$, $p \in \{0.1, 0.01, 0.001\}$) for $\lfloor nT \rfloor$ realizations of (ARMA), (ARCH) and (SV)

investigated under the null:

$$X_{i,n} = \begin{cases} 0.3 \cdot X_{i-1} + Z_i + 0.7 \cdot Z_{i-1}, & i = 1, \ldots, \lfloor nt^* \rfloor, \\ 0.8 \cdot X_{i-1} + Z_i + 0.7 \cdot Z_{i-1}, & i = \lfloor nt^* \rfloor + 1, \ldots, \lfloor nT \rfloor, \end{cases} \quad Z_i \overset{\text{i.i.d.}}{\sim} t_{10}, \tag{4.15}$$

$$X_{i,n}^2 = \begin{cases} (0.01 + 0.3125 \cdot X_{i-1}^2) Z_i^2, & i = 1, \ldots, \lfloor nt^* \rfloor, \\ (0.01 + 0.5773 \cdot X_{i-1}^2) Z_i^2, & i = \lfloor nt^* \rfloor + 1, \ldots, \lfloor nT \rfloor, \end{cases} \quad Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{4.16}$$

$$|X_i| = \sigma_i |Z_{i,n}|, \quad \text{where } Z_{i,n} \overset{\text{i.ni.d.}}{\sim} \begin{cases} t_{0.5}, & i = 1, \ldots, \lfloor nt^* \rfloor, \\ t_1, & i = \lfloor nt^* \rfloor + 1, \ldots, \lfloor nT \rfloor, \end{cases} \tag{4.17}$$

where $n = 125, 500$ and $T = 4$ as before and $t^* = 1.15, 2.5$, corresponding to breaks after 5% and 50% of the observation period. In the ARMA(1,1)-model with the break in the AR-parameter from 0.3 to 0.8 there is no break in the tail index, but a break in the variance from 0.92 to 1.81, i.e., a more volatile distribution after the break. In the ARCH case the parameter shift induces a tail index change from $8/2 = 4$ to $4/2 = 2$ (cf. Davis and Mikosch, 1998, Table 1), i.e., heavier tails after the break. At the same time the variance is finite pre-break and (hairline) infinite post-break. For the stochastic volatility model with the break in the error distribution the break is in the opposite direction with a change in the tail index from 0.5 to 1.

Note that for the ARMA(1,1) model in (4.15) the null hypothesis $\mathcal{H}_{0,\gamma}$ is true. However, as in finite samples an upward break in the variance may not be clearly discerned from one in the tail index by our procedure, we should expect more rejections for $\tau_n^{V/W^{\widehat{\gamma}}}$ than in Table 4.2. This is generally confirmed by the results in Table 4.3. Furthermore, the variance change is most frequently detected using $\tau_n^{V/W^{\widehat{x}_p}}$ for $p = 0.1$. This may be explained by the higher variance of the estimates $\widehat{x}_p$ for smaller values of p, which makes detection of a quantile break very difficult, if the quantiles do not lie sufficiently far apart, as is plausible here, where a mere variance change occurred.

If there is a break in the tail index and the variance as in the ARCH- and SV-case, one can see from Table 4.3 that the procedures based on the Weissman (1978) estimator clearly perform better than those based solely on the Hill (1975) estimator. Heuristically, this may be explained by the fact that the Weissman (1978) estimator also takes differences in scale before and after the break into account (via $X_k(s,t,1)$; see (4.5)). Since in reality, changes in the tail index will most likely result in changes of scale, one should use the tests based on $V/W^{\widehat{x}_p}$. Further, the choice $\tau_n^{V^{\widehat{x}_p}}$ with $p = 0.1$ leads to the highest power, particularly for small sample lengths and the

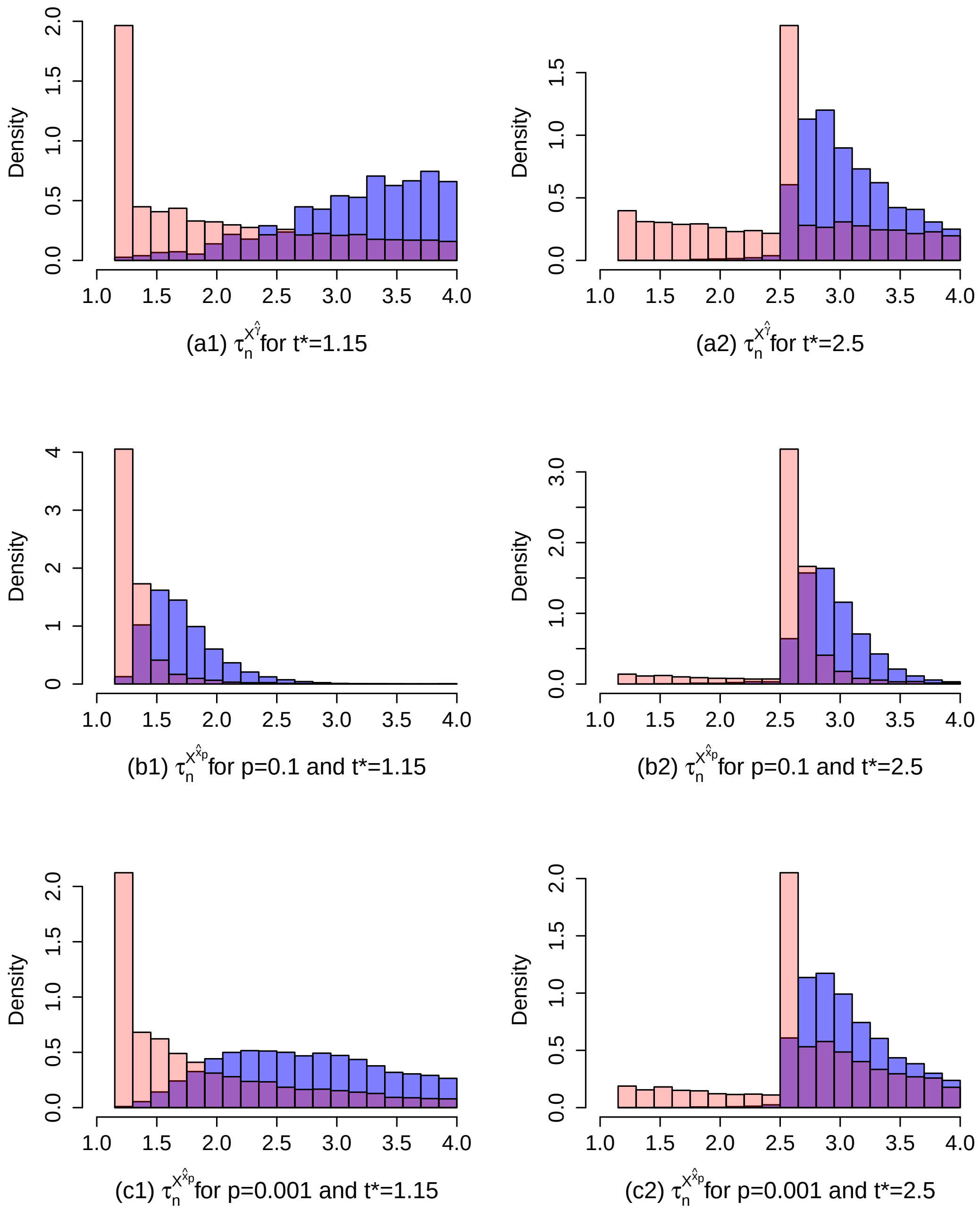

Figure 4.1: Histograms of detection times $\tau_n^{X^{\widehat{\gamma}}}$, $\tau_n^{X^{\widehat{x_p}}}$ for $X = V$ (bright blue) and for $X = W$ (bright red) for $p = 0.1, 0.001$ for (4.15) and $t^* = 1.15$ ((a1), (b1), (c1)), $t^* = 2.5$ ((a2), (b2), (c2)) for $\lfloor nT \rfloor = 2000$

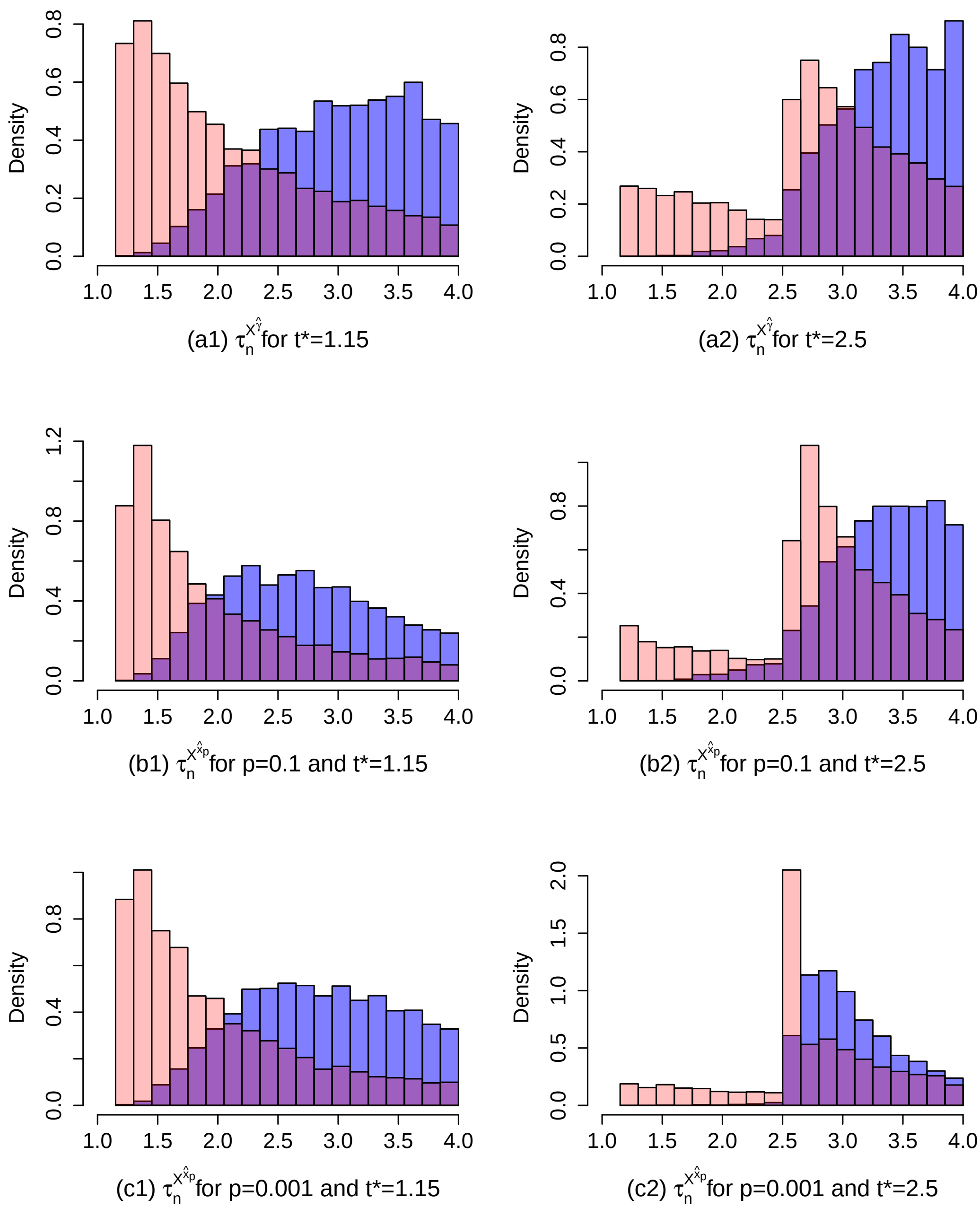

Figure 4.2: Histogram of detection times $\tau_n^{X\widehat{\gamma}}$, $\tau_n^{X\widehat{x_p}}$ for $X = V$ (bright blue) and for $X = W$ (bright red) for $p = 0.1, 0.001$ for (4.16) and $t^* = 1.15$ ((a1), (b1), (c1)), $t^* = 2.5$ ((a2), (b2), (c2)) for $\lfloor nT \rfloor = 2000$

downward break in tail heaviness for model (4.17).

Recall that the procedure based on $V^{\widehat{\gamma}}$ is inconsistent under $\mathcal{H}_{1,\gamma}^{>}$. While this is not yet apparent for the early break ($t^* = 1.15$), it is for the late break ($t^* = 2.5$), where power is significantly lower as a larger portion of the sample upon which $\widehat{\gamma}(1,t)$ is based is 'contaminated' by very heavy tailed observations.

For sequential tests like ours, power is not the only criterion by which to judge a procedure, but also how promptly changes are detected. To look into this, Figures 4.1 and 4.2 show histograms of the (finite) realizations of $\tau_n^{V/W^{\widehat{\gamma}}}$ and $\tau_n^{V/W^{\widehat{x_p}}}$ (bright blue / red) at the 10% level for the ARMA and the ARCH models given in (4.15) and (4.16) respectively with $\lfloor nT \rfloor = 2000$. The histogram for (4.17) does not provide any additional insights and is omitted. There are 19 bars in all plots with breaks at $1 + l \cdot 0.15$ ($l = 1, \ldots, 20$). The value of t^* at which the changes are located are given by $1.15 = 1 + 1 \cdot 0.15$ ($l = 1$) and $2.5 = 1 + 10 \cdot 0.15$ ($l = 10$).

The results for the ARMA model are displayed in Figure 4.1. The high false detection rate for the tail index-based method using the detector $W_n^{\widehat{\gamma}}$ seems largely to be due to false detections just shortly after the break, as can be seen from the distinctive peaks in panels (a1) and (a2). The detections with $\tau_n^{W^{\widehat{x_p}}}$ for $p = 0.1$ in (b1) and (b2) indicate that a very large portion of detections occur within the time corresponding to the two bars right after the break. This holds to a lesser extent for the results shown in (c1) and (c2), where, however, detection rates were poor. Overall the detection speed is satisfactory but faster for larger values of p. For the ARCH model one can see slightly dissimilar detection patterns for all procedures based on W. The highest number of detections always occurs one or two bars after the break and that rate goes down only slowly so that detections (if they occur) take on average longer than in the ARMA case. This may be explained by the fact that ARCH models incorporate conditional heteroscedasticity, such that detection of changes in the variability of time series is inherently more difficult. We need to observe longer periods of higher volatility before one can reject the null here.

Comparing these results with those for the procedures based on V we see that for the latter detections take much longer. They never peak in the initial period, where the change occurs. This introduces a delicate trade-off for the detectors we introduced. The stopping times based on V terminate more often under the alternative than those based on W, but they take longer to do so. So if a swift detection is of the utmost importance, we recommend to use $\tau_n^{W^{\widehat{x_p}}}$ for $p = 0.01$. If it is more important that a break is detected at all, but speed is of lesser interest, then $\tau_n^{V^{\widehat{x_p}}}$ for $p = 0.01$ seems to be the wisest choice, unless a break towards lighter tails is expected in which case power was rather dismal.

4.4 Application

In this section we apply our tests to the lower tail of log-returns, i.e., log-losses, of Bank of America stocks covering the period of the financial crisis of 2007-2008, where short selling US financial stocks was halted until October 2, 2008 following an SEC order released on September 19, 2008. The return series we consider is displayed in the top part of Figure 4.4. Results for stock prices of other US financial companies (Morgan Stanley, Citigroup and Goldman Sachs) are very similar and can be obtained upon request. Since we try to detect changes in unconditional quantiles, our focus is on the long-term distributional changes in the time series, not on short-term changes in the conditional distribution. We set our (artificial) training and observation period to be the years from 2005 to 2012 corresponding to 2013 observations, $X_1, \dots, X_{2012}$, the first quarter of which (roughly the years 2005 and 2006) make up the training period. The lengths of the training and observation period were chosen to correspond to the case $n = 500$ in the simulations, for which size and power proved to be very satisfying. Furthermore, we choose the training period to precede the onset of the financial crisis in 2007, so that we may analyze the performance of our procedures during these tumultuous years.

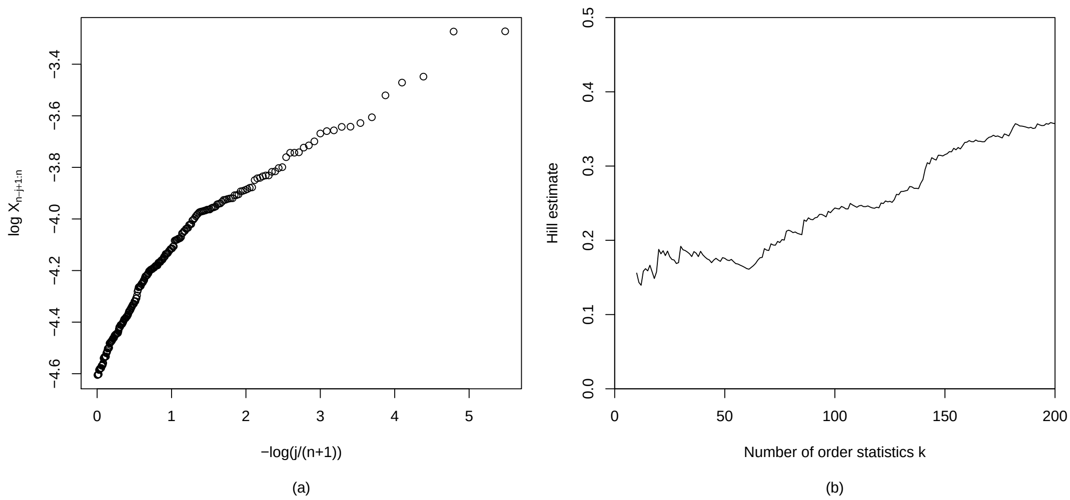

Figure 4.3: (a) Pareto quantile plot of shifted data. (b) Hill estimates as a function of upper order statistics k used in the estimation.

Given the very calm behavior of the log-returns during the training period one may have suspected that a break toward heavier tails is much more likely than one toward lighter tails. Additionally, it is vital for managing risk adequately to detect a break in the tail behavior quickly, because if it is registered to late the cost of hedging that risk may already have increased dramatically. For these two reasons we focus on the detectors W, which performed only slightly worse than the detectors V when there is break leading to heavier tails, yet detected those much faster.

Next, we verify that our two central assumptions, the stationary mixing assumption **(C2)** and the heavy tail assumption **(C5)**, are plausibly met by the time series in the training period. To check whether there is evidence for heavy tails we plot the Pareto quantile plot in Figure 4.3 (a), where the points $(-\log(j/(n+1)), \log X_{n-j+1:n})$, $j = 1, \ldots, n$, are plotted. See Beirlant *et al.* (1996) for more on Pareto quantile plots. In order for all $\log X_{n-j+1:n}$'s to be well-defined we shifted the observations to the positive half-line by adding the absolute value of the smallest return plus 0.01. An upward sloping linear trend, like the one that can be seen in Figure 4.3 (a) from $-\log(j/(n+1)) \approx 1.5$ onwards, for some $j = 1, \ldots, k+1$ in the plot indicates a good fit of the tail to (strict) Pareto behavior. An estimate of $1/\alpha$ can then be obtained as the slope of the line from the point $(-\log((k+1)/(n+1)), \log X_{n-k:n})$ onwards, where the slope seems to be roughly 0.2. This is confirmed in the (slightly upward trending) Hill plot in Figure 4.3 (b), which displays the Hill estimates of the shifted data as a function of the upper order statistics k used in the estimation. As for the mixing assumption **(C2)** the best ARCH(p)-model (by AIC) was an ARCH(1). Using an ARCH-LM test however, we could not reject the null of no conditional heteroscedasticity for this model (p-value of 0.86). Routine testing and plotting of the autocorrelation function of the raw and squared log-losses also revealed no dependence in the data, such that the data may reasonably be regarded as independent. Further, applying the retrospective tests of Chapter 2 and 3 we found no evidence of extreme quantile or tail index breaks during that period which would violate the stationarity assumption. Hence, we proceed with our monitoring procedure.

Since their inception by Engle (1982) (G)ARCH-models have arguably become the most popular models for returns on risky assets. So the absence of ARCH-effects in the training period may be surprising. However, Stărică and Granger (2005) argue for models of returns that are locally i.i.d. In our case the period from 2005 to 2006 seems to a period, where returns behave like an i.i.d. sequence.

The results are shown in the middle part of Figure 4.4. As in the simulations we choose $k/n = 0.2$ and $t_0 = 0.2$. All procedures terminate at the 5%-level if the value of 45.4 is exceeded by the detector. We see that the procedure testing for a change in the 10%-quantile of the log-returns terminates first in November of 2007, followed later by the detection of a break in the 1%-quantile in August 2008. A 0.1%-quantile break is detected last in early 2009. However, we find no evidence of a tail index break in the observation period. The lower part of Figure 4.4 sheds some light on this phenomenon. There, the rolling window extreme value index estimates based on samples of length $\lfloor nt_0 \rfloor = 100$, that the detectors are based on, are presented. The

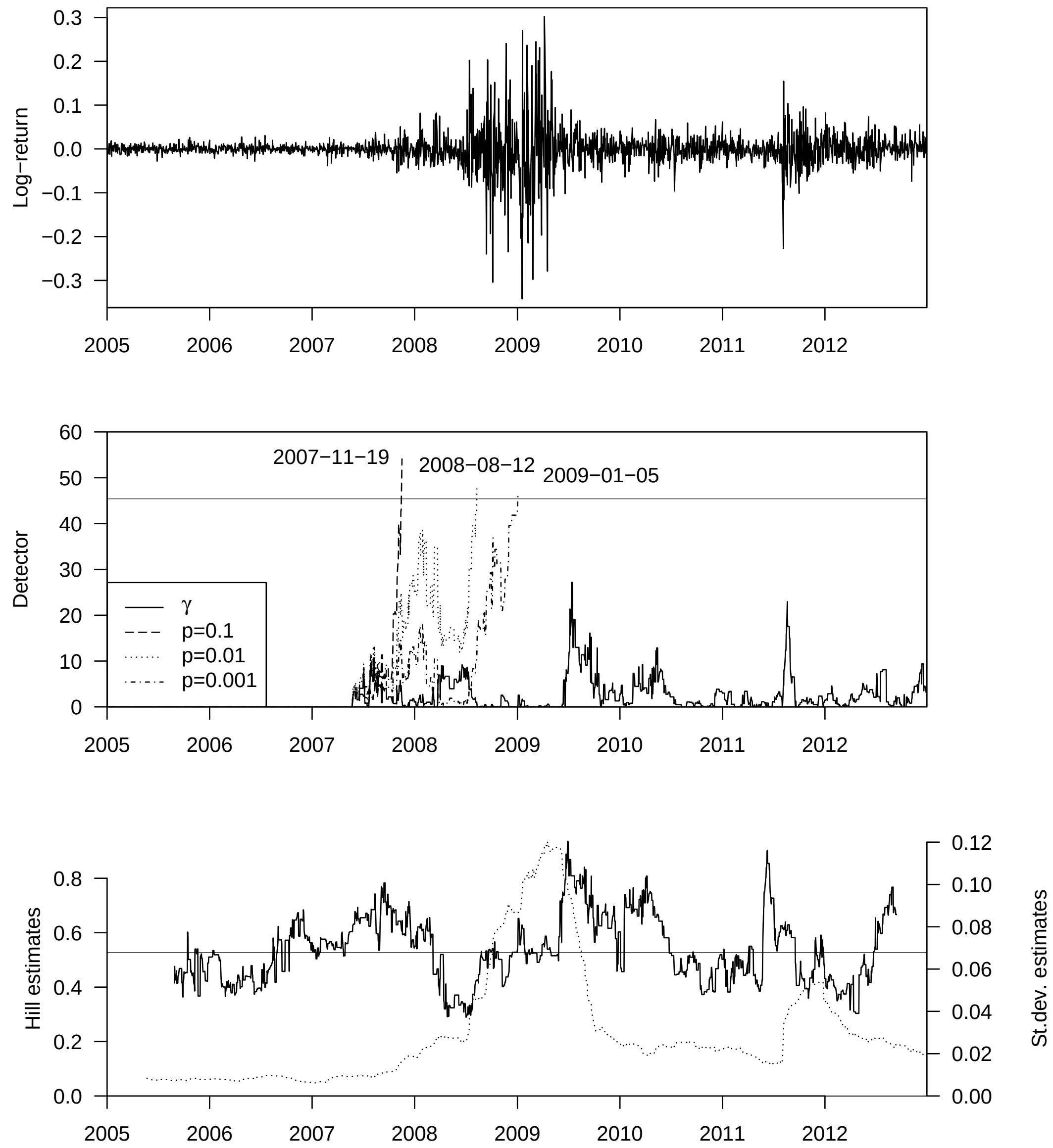

Figure 4.4: Top panel: Log-returns of Bank of America stock. Middle panel: Values of detectors $\widehat{W^{\gamma}}$ (solid), and $\widehat{W^{x_p}}$ for $p = 0.1, 0.001, 0.0001$ (dashed, dotted, dash-dotted) and value of 5%-threshold (horizontal solid line). Bottom panel: Rolling Hill estimate (jagged solid line), Hill estimate based on training period (straight solid line), and standard deviation estimates (dotted line).

estimates hover around the value of 0.6 during the whole period, which is roughly the extreme value index estimate of 0.52 based on the training period. (Due to the location dependence of Hill estimates, the extreme value index in the training period was estimated to be roughly 0.2 for the shifted data in Figure 4.3 (b) and 0.52 now, based on non-shifted returns. The extreme value index itself is of course shift-invariant.) This contrasts with the behavior of the standard deviation estimates based on the same rolling windows, where we see a marked spike peaking in early 2009. Hence, we find indications that the change in the extreme quantiles is not caused by a change in the tail index but rather by a change in the scale of the log-returns. Largely, the above results are consistent with the simulations under the alternative, where a variance change occured. Procedures based on $W^{\widehat{x_p}}$ detect mere variance changes more easily for larger values of p, while that based on $W^{\widehat{\gamma}}$ did not pick up a tail index change.

4.5 Proofs

Proof of Theorem 4.1: The proof of *(i)* mainly rests on a time shifted version of the (weighted) weak convergence established in Theorem 2.5, see also the proof of Theorem 2.1. That is, for some $\delta > 0$ and any $\nu \in [0, 1/2)$,

$$\frac{\sqrt{k}}{y^\nu}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}}-\frac{k}{n}y\right]\right\}\xrightarrow[(n\to\infty)]{\mathcal{D}}\frac{1}{y^\nu}W(t,y)\quad\text{in } D\left([t_0,T]\times[0,y_0+\delta]\right) \tag{4.18}$$

for a sequence of uniformly distributed random variables $U_i \sim \mathcal{U}[0,1]$ satisfying **(C2)**-**(C4)**, where $W(t,y)$ is a continuous centered Gaussian process with covariance function

$$\operatorname{Cov}\left(W\left(t_1,y_1\right),W\left(t_2,y_2\right)\right)=\min\left(t_1,t_2\right)r\left(y_1,y_2\right)$$

and $r(\cdot,\cdot)$ defined in **(C3)**. Then for X_i satsifying **(C2)-(C5)**, by the proof of Theorem 3.1 in Drees (2000), the uniformly distributed $U_i := F(X_i)$ satisfy **(C2)-(C4)** and

$$X_i > U\left(\frac{n}{ky}\right)\quad\Longleftrightarrow\quad U_i>1-\frac{k}{n}y.$$

Hence,

$$\frac{\sqrt{k}}{y^\nu}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}I_{\left\{X_i>U\left(\frac{n}{ky}\right)\right\}}-ty\right\}\xrightarrow[(n\to\infty)]{\mathcal{D}}\frac{1}{y^\nu}W(t,y)\quad\text{in } D\left([t_0,T]\times[0,y_0+\delta]\right).$$

The continuous mapping theorem (CMT) then implies

$$\frac{\sqrt{k}}{y^{\nu}}\begin{pmatrix}\frac{1}{k}\sum_{i=n+1}^{\lfloor n\max(1+t_0,t)\rfloor} I_{\left\{X_i>U(\frac{n}{ky})\right\}}-(\max(1+t_0,t)-1)y\\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor+1}^{\lfloor nt\rfloor} I_{\left\{X_i>U(\frac{n}{ky})\right\}}-t_0y\\ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>U(\frac{n}{ky})\right\}}-ty\end{pmatrix}$$
$$\xrightarrow[(n\to\infty)]{\mathcal{D}}\frac{1}{y^{\nu}}\begin{pmatrix}W(\max(1+t_0,t),y)-W(1,y)\\ W(t,y)-W(t-t_0,y)\\ W(t,y)\end{pmatrix}\quad\text{in } D^3\left([t_0,T]\times[0,y_0+\delta]\right).$$

Now invoking a Skorohod construction (e.g., Wichura, 1970, Thm. 1) we get on a suitable probability space

$$\sup_{\substack{t\in[t_0,T]\\ y(0,y_0+\delta]}}\frac{1}{y^{\nu}}\left|\sqrt{k}\begin{pmatrix}\frac{1}{k}\sum_{i=n+1}^{\lfloor n\max(1+t_0,t)\rfloor} I_{\left\{X_i>U(\frac{n}{ky})\right\}}-(\max(1+t_0,t)-1)y\\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor+1}^{\lfloor nt\rfloor} I_{\left\{X_i>U(\frac{n}{ky})\right\}}-t_0y\\ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>U(\frac{n}{ky})\right\}}-ty\end{pmatrix}\right.$$
$$\left.-\begin{pmatrix}W(\max(1+t_0,t),y)-W(1,y)\\ W(t,y)-W(t-t_0,y)\\ W(t,y)\end{pmatrix}\right|\xrightarrow[(n\to\infty)]{\text{a.s.}}0.$$

Arguing componentwise as in the proof of Einmahl *et al.* (2016, Cor. 3) this implies for some $\widetilde{\delta}>0$ (using the second-order condition **(C5)**)

$$\sup_{\substack{t\in[t_0,T]\\ y\geq y_0^{-1/\gamma}-\widetilde{\delta}}} y^{\nu/\gamma}\left|\sqrt{k}\begin{pmatrix}\frac{1}{k}\sum_{i=n+1}^{\lfloor n\max(1+t_0,t)\rfloor} I_{\left\{X_i>yU\left(\frac{n}{k}\right)\right\}}-(\max(1+t_0,t)-1)y^{-1/\gamma}\\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor+1}^{\lfloor nt\rfloor} I_{\left\{X_i>yU\left(\frac{n}{k}\right)\right\}}-t_0y^{-1/\gamma}\\ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>yU\left(\frac{n}{k}\right)\right\}}-ty^{-1/\gamma}\end{pmatrix}\right.$$
$$\left.-\begin{pmatrix}W(\max(1+t_0,t),y^{-1/\gamma})-W(1,y^{-1/\gamma})\\ W(t,y^{-1/\gamma})-W(t-t_0,y^{-1/\gamma})\\ W(t,y^{-1/\gamma})\end{pmatrix}\right|\xrightarrow[(n\to\infty)]{\text{a.s.}}0,\quad(4.19)$$

where Proposition 2.2 justifies the final step of their proof in our case. Now retrace the proof of Corollary 2.1 to see that

$$\sup_{\substack{t\in[t_0,T]\\ y\geq y_0^{-1/\gamma}}} y^{\nu/\gamma}\left|\sqrt{k}\begin{pmatrix}\frac{1}{k}\sum_{i=n+1}^{\lfloor n(1+t_0)\vee t\rfloor} I_{\{X_i>yX_k(1,(1+t_0)\vee t,1)\}}-((1+t_0)\vee t-1)y^{-1/\gamma}\\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor+1}^{\lfloor nt\rfloor} I_{\{X_i>yX_k(t-t_0,t,1)\}}-t_0y^{-1/\gamma}\\ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i>yX_k(0,t,1)\}}-ty^{-1/\gamma}\end{pmatrix}\right.$$
$$\left.-\begin{pmatrix}W((1+t_0)\vee t,y^{-1/\gamma})-W(1,y^{-1/\gamma})-y^{-1/\gamma}\left[W((1+t_0)\vee t,1)-W(1,1)\right]\\ W(t,y^{-1/\gamma})-W(t-t_0,y^{-1/\gamma})-y^{-1/\gamma}\left[W(t,1)-W(t-t_0,1)\right]\\ W(t,y^{-1/\gamma})-y^{-1/\gamma}W(t,1)\end{pmatrix}\right|$$
$$\xrightarrow[(n\to\infty)]{\text{a.s.}} 0. \tag{4.20}$$

The key ingredient in this step justifying the replacement of $U(n/k)$ in (4.19) by the respective $X_k(s,t,1)$ in (4.20) is the generalized Vervaat lemma in Einmahl *et al.* (2010, Lem. 5), which gives a weak convergence result for $X_k(0,t,1)/U(n/k)$ from the third component of (4.19) (and similarly for the other components). The convergence in (4.20) is the key result with which one may deduce weak convergence results for various tail index estimators, see Examples 2.3-2.5. As already mentioned, we focus here on the Hill estimator $\widehat{\gamma}:=\widehat{\gamma}_H$ defined in (4.6). Focus on the third component of (4.20) (the others can be dealt with similarly) and notice that for $j=0,1,\ldots,\lfloor kt\rfloor-1$

$$\frac{1}{\lfloor kt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i>yX_k(0,t,1)\}}=\begin{cases}\frac{\lfloor kt\rfloor-j}{\lfloor kt\rfloor}, & y\in\left[\frac{X_{\lfloor nt\rfloor-\lfloor kt\rfloor+j:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor}},\frac{X_{\lfloor nt\rfloor-\lfloor kt\rfloor+j+1:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor}}\right),\\ 0, & y\geq\frac{X_{\lfloor nt\rfloor:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor}}.\end{cases}$$

Then it is an easy exercise in integration to check that

$$\sqrt{k}\left(\widehat{\gamma}_H(0,t)-\gamma\right)=\sqrt{k}\left(\int_1^\infty\frac{1}{\lfloor kt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i>yX_k(0,t,1)\}}-y^{-1/\gamma}\frac{\mathrm{d}y}{y}\right).$$

Similar representations hold for $\widehat{\gamma}(1,\max(1+t_0,t))$ and $\widehat{\gamma}(t-t_0,t)$, so that (4.20) implies

$$\sqrt{k}\begin{pmatrix}(\max(1+t_0,t)-1)\left(\widehat{\gamma}(1,\max(1+t_0,t))-\gamma\right)\\ t_0\left(\widehat{\gamma}(t-t_0,t)-\gamma\right)\\ t\left(\widehat{\gamma}(0,t)-\gamma\right)\end{pmatrix}\xrightarrow[(n\to\infty)]{\mathcal{D}}$$

$$\sigma_{\widehat{\gamma},\gamma} \begin{pmatrix} W(\max(1+t_0,t)) - W(1) \\ W(t) - W(t-t_0) \\ W(t) \end{pmatrix} \quad \text{in } D^3[t_0,T], \quad (4.21)$$

where $\sigma^2_{\widehat{\gamma},\gamma} = \gamma^2 \int_0^1 \int_0^1 \left\{ \frac{r(x,y)}{xy} - \frac{r(x,1)}{x} - \frac{r(1,y)}{y} + r(x,y) \right\} \mathrm{d}x\mathrm{d}y$ and $W(\cdot)$ a standard Brownian motion. Notice for this that by calculating covariances

$$\int_1^\infty \left[W(t,y^{-1/\gamma}) - y^{-1/\gamma} W(t,1) \right] \frac{\mathrm{d}y}{y} = \gamma \int_0^1 \left[W(t,u) - uW(t,1) \right] \frac{\mathrm{d}u}{u} \stackrel{\mathcal{D}}{=} \sigma_{\widehat{\gamma},\gamma} W(t).$$

From (4.21) and the CMT we obtain

$$V_n^{\widehat{\gamma}}(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{W(t) - tW(1)}{\int_{t_0}^1 \left(W(s) - sW(1)\right)^2 \mathrm{d}s} \quad \text{in } D[1+t_0,T], \quad (4.22)$$

$$W_n^{\widehat{\gamma}}(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{W(t) - W(t-t_0) - t_0 W(1)}{\int_{t_0}^1 \left(W(s) - W(s-t_0) - t_0 W(1)\right)^2 \mathrm{d}s} \quad \text{in } D[1+t_0,T].$$

The result is now proved via another application of the CMT.

For part *(ii)* we observe that it follows from (4.21) similarly as in the proof of Theorem 3.1 that (for the general idea of how to derive convergence of $\widehat{x}_p$ from that of $\widehat{\gamma}$ see (4.27) below and the steps following it, in particular (4.29)).

$$\frac{\sqrt{k}}{\log\left(k/(np)\right)} \begin{pmatrix} (\max(1+t_0,t)-1)\log\left(\frac{\widehat{x}_p(1,\max(1+t_0,t))}{U(1/p)}\right) \\ t_0 \log\left(\frac{\widehat{x}_p(t-t_0,t)}{U(1/p)}\right) \\ t\log\left(\frac{\widehat{x}_p(0,t)}{U(1/p)}\right) \end{pmatrix}$$

$$\xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma},\gamma} \begin{pmatrix} W(\max(1+t_0,t)) - W(1) \\ W(t) - W(t-t_0) \\ W(1) \end{pmatrix} \quad \text{in } D^3[t_0,T], \quad (4.23)$$

whence with the CMT again

$$V_n^{\widehat{x}_p}(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{W(t) - tW(1)}{\int_{t_0}^1 \left(W(s) - sW(1)\right)^2 \mathrm{d}s} \quad \text{in } D[1+t_0,T],$$

$$W_n^{\widehat{x}_p}(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{W(t) - W(t-t_0) - t_0 W(1)}{\int_{t_0}^1 \left(W(s) - W(s-t_0) - t_0 W(1)\right)^2 \mathrm{d}s} \quad \text{in } D[1+t_0,T].$$

The conclusion follows as before. □

Proof of Theorem 4.2: We first prove the two statements in *(i)*. Under $\mathcal{H}_{1,\gamma}^{>}$ we get by an adaptation of the proof of Theorem 2.3 similar to the one in the proof of Theorem 4.1

$$\sqrt{k}\left(\widehat{\gamma}(1,t)-\gamma_{\text{pre}}\right) \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \frac{B_{\text{pre}}(t)-B_{\text{pre}}(1)}{t-1} \quad \text{in } D\left[1+t_0,T\right] \tag{4.24}$$

and by a further close inspection that even

$$\sqrt{k}\begin{pmatrix}\widehat{\gamma}(1,\max(1+t_0,t))-\gamma_{\text{pre}}\\ \widehat{\gamma}(0,t)-\gamma_{\text{pre}}\end{pmatrix} \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \begin{pmatrix}(B_{\text{pre}}(\max(1+t_0,t))-B_{\text{pre}}(1))/(\max(1+t_0,t)-1)\\ B_{\text{pre}}(t)/t\end{pmatrix}$$

jointly in $D^2[t_0,T]$, where, setting $t_{\min}=\min(t,t^*)$,

$$B_{\text{pre}}(t)=\gamma_{\text{pre}}\int_0^1\left[W_{\text{pre}}(t_{\min},u\frac{t}{t_{\min}})-uW_{\text{pre}}(t_{\min},\frac{t}{t_{\min}})\right]\frac{\mathrm{d}u}{u}$$

and $W_{\text{pre}}(\cdot,\cdot)$ is as $W(\cdot,\cdot)$ in (4.18) with $r(\cdot,\cdot)$ replaced by $r_{\text{pre}}(\cdot,\cdot)$. Hence, (4.22) holds with $W(\cdot)$ replaced by $B_{\text{pre}}(\cdot)$, where $B_{\text{pre}}(\cdot)$ is also a continuous centered Gaussian process. The result follows.

For the other part of *(i)* it suffices to note that

$$\widehat{\gamma}(0,1) \underset{(n\to\infty)}{\xrightarrow{P}} \gamma_{\text{pre}}<\gamma_{\text{post}} \underset{(n\to\infty)}{\xleftarrow{P}} \widehat{\gamma}(1,T),$$

because by (4.24) $\widehat{\gamma}$ will converge in probability to the dominant tail index (i.e., $\max(\gamma_{\text{pre}},\gamma_{\text{post}})$) in a sample with one tail index break.

For *(ii)* simply note that

$$\widehat{\gamma}(0,1) \underset{(n\to\infty)}{\xrightarrow{P}} \gamma_{\text{pre}}<\gamma_{\text{post}} \underset{(n\to\infty)}{\xleftarrow{P}} \widehat{\gamma}(t^*,t^*+t_0).$$

If $\gamma_{\text{pre}}>\gamma_{\text{post}}$ in *(iii)* (the other case is similar) we can deduce from (4.23) that

$$\frac{\sqrt{k}}{\log\left(k/(np)\right)}\log\left(\frac{\widehat{x}_p(0,1)}{U_{\text{pre}}(1/p)}\right)=\mathcal{O}_{\mathrm{P}}(1).$$

If we also had

$$\frac{\sqrt{k}}{\log\left(k/(np)\right)}\log\left(\frac{\widehat{x}_p(1,T)}{U_{\text{pre}}(1/p)}\left(\frac{T}{t^*}\right)^{\gamma_{\text{pre}}}\right)=\mathcal{O}_{\mathrm{P}}(1), \tag{4.25}$$

the result would follow, because

$$V_n^{\widehat{x}_p}(T) = \frac{\left[(t-1)\frac{\sqrt{k}}{\log(k/(np))}\left\{\log\left(\frac{\widehat{x}_p(1,T)}{U_{\text{pre}}(1/p)}\right) - \log\left(\frac{\widehat{x}_p(0,1)}{U_{\text{pre}}(1/p)}\right)\right\}\right]^2}{\int_{t_0}^1 \left[s\frac{\sqrt{k}}{\log(k/(np))}\log\left(\frac{\widehat{x}_p(0,s)}{\widehat{x}_p(0,1)}\right)\right]^2 \mathrm{d}s}. \tag{4.26}$$

To show (4.25) decompose

$$\begin{aligned}\left(\frac{t}{t_{\min}}\right)^{\gamma_{\text{pre}}}\widehat{x}_p(1,t) - U_{\text{pre}}\left(\frac{1}{p}\right) =& \left[\left(\frac{t}{t_{\min}}\right)^{\gamma_{\text{pre}}} X_k(1,t,1) - U_{\text{pre}}\left(\frac{n}{k}\right)\right]\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}\\
&+\left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)} - \left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}\right]U_{\text{pre}}\left(\frac{n}{k}\right)\\
&+\left[U_{\text{pre}}\left(\frac{n}{k}\right)\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}} - U_{\text{pre}}\left(\frac{1}{p}\right)\right]\\
=& I + II + III. \end{aligned} \tag{4.27}$$

Before considering these three terms separately, observe that by the mean value theorem, using $\frac{\partial}{\partial\tau}(x^\tau) = x^\tau\log(x)$, there exists a $\xi \in [-1,1]$ such that

$$\begin{aligned}&\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)} - \left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}\\
&= \left(-\widehat{\gamma}(1,t) + \gamma_{\text{pre}}\right)\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}+\xi\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)}\log\left(\frac{np}{k}\right)\\
&= \left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}\left(\gamma_{\text{pre}} - \widehat{\gamma}(1,t)\right)\left(\frac{np}{k}\right)^{\xi\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)}\log\left(\frac{np}{k}\right).\end{aligned}$$

Then use

$$\begin{aligned}\left(\frac{np}{k}\right)^{\xi\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)} &= \exp\left[\xi\left(\gamma_{\text{pre}} - \widehat{\gamma}(1,t)\right)\log\left(\frac{np}{k}\right)\right]\\
&= \exp\left[\xi\mathcal{O}_{\mathrm{P}}\left(\frac{1}{\sqrt{k}}\right)\log\left(\frac{np}{k}\right)\right] \xrightarrow[(n\to\infty)]{\mathrm{P}} 1\end{aligned}$$

uniformly in t (by **(C6)** and (4.24)) and **(C7)** to get

$$
\begin{aligned}
&\frac{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}-\left(\frac{np}{k}\right)^{-\gamma_{\mathrm{pre}}}\right] && (4.28)\\
=&\frac{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\gamma_{\mathrm{pre}}}\left(\gamma_{\mathrm{pre}}-\widehat{\gamma}(1,t)\right)\left(\frac{np}{k}\right)^{\xi\left(\gamma_{\mathrm{pre}}-\widehat{\gamma}(1,t)\right)}\log\left(\frac{np}{k}\right)\right]\\
=&\left(1+o_{\mathrm{P}}(1)\right)\left(\gamma_{\mathrm{pre}}-\widehat{\gamma}(1,t)\right)\log\left(\frac{np}{k}\right). && (4.29)
\end{aligned}
$$

Furthermore applying the functional delta method (e.g., van der Vaart and Wellner, 1996, Theorem 3.9.4) to (a slight adaptation of) Eq. 2.69 yields

$$
\begin{aligned}
&\sqrt{k}\left\{\frac{X_k(1,t,1)}{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}-\left(\frac{t}{t_{\min}}\right)^{-\gamma_{\mathrm{pre}}}\right\}\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}\\
&-\gamma_{\mathrm{pre}}\left(\frac{t}{t_{\min}}\right)^{-(\gamma_{\mathrm{pre}}+1)}\left[\frac{W_{\mathrm{pre}}\left(t_{\min},\frac{t}{t_{\min}}\right)}{t_{\min}}-W_{\mathrm{pre}}\left(1,\frac{t}{t_{\min}}\right)\right]\quad\text{in } D[1+t_0,T],
\end{aligned}
\tag{4.30}
$$

where we used that

$$
\phi: D[1+t_0,T]\to D[1+t_0,T],\quad \phi(f(\cdot))=f^{-\gamma_{\mathrm{pre}}}(\cdot)
$$

is Hadamard-differentiable tangentially to $C[1+t_0,T]$ in $t\mapsto t/t_{\min}$ with derivative

$$
\phi'_{t/t_{\min}}(f(\cdot))=-\gamma_{\mathrm{pre}}\left(\frac{t}{t_{\min}}\right)^{-(\gamma_{\mathrm{pre}}+1)}f(\cdot).
$$

Using this result in the fourth equality we get

$$
\begin{aligned}
&\frac{I}{U_{\mathrm{pre}}(1/p)\log\left(\frac{k}{np}\right)}\\
=\;&\frac{1}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)}\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}\frac{1}{\log\left(\frac{k}{np}\right)}\left[X_k(1,t,1)-U_{\mathrm{pre}}\left(\frac{n}{k}\right)\right]\\
=\;&\frac{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}-\left(\frac{np}{k}\right)^{-\gamma_{\mathrm{pre}}}+\left(\frac{np}{k}\right)^{-\gamma_{\mathrm{pre}}}\right]\frac{\left[\frac{X_k(0,t,1)}{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}-1\right]}{\log\left(\frac{k}{np}\right)}
\end{aligned}
$$

$$
\begin{aligned}
&\underset{(4.29)}{\overset{\textbf{(C7)}}{=}} \left[(1+o_{\mathrm{P}}(1))(\gamma_{\mathrm{pre}}-\widehat{\gamma}(1,t))\log\left(\frac{np}{k}\right)+1+o\left(\frac{1}{\sqrt{k}}\right)\right]\frac{\left[\frac{X_k(0,t,1)}{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}-1\right]}{\log\left(\frac{k}{np}\right)}\\
&\overset{(4.30)}{=} \left[(1+o_{\mathrm{P}}(1))\sqrt{k}(\gamma_{\mathrm{pre}}-\widehat{\gamma}(1,t))\frac{\log\left(\frac{np}{k}\right)}{\sqrt{k}}+1+o\left(\frac{1}{\sqrt{k}}\right)\right]\frac{\mathcal{O}_P\left(\frac{1}{\sqrt{k}}\right)}{\log\left(\frac{k}{np}\right)}\\
&\underset{(4.24)}{\overset{\textbf{(C6)}}{=}} o_{\mathrm{P}}\left(1/\sqrt{k}\right)
\end{aligned}
$$

uniformly in t. Further, utilizing **(C6)** and **(C7)** for the third term,

$$
\frac{III}{U_{\mathrm{pre}}(1/p)\log\left(\frac{k}{np}\right)}=\frac{1}{\log\left(\frac{k}{np}\right)}\left[\frac{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)\left(\frac{np}{k}\right)^{-\gamma_{\mathrm{pre}}}}-1\right]=o\left(1/\sqrt{k}\right).
$$

Using (4.29) and (4.24) we get for the last term

$$
(t-1)\sqrt{k}\frac{II}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)\log\left(\frac{k}{np}\right)}\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} B_{\mathrm{pre}}(t)-B_{\mathrm{pre}}(1)\quad\text{in } D\left[1+t_0,T\right],
$$

whence

$$
\frac{(t-1)\sqrt{k}}{\log\left(\frac{k}{np}\right)}\left\{\frac{\left(\frac{t}{t_{\min}}\right)^{\gamma_{\mathrm{pre}}}\widehat{x}_p(1,t)}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)}-1\right\}\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} B_{\mathrm{pre}}(t)-B_{\mathrm{pre}}(1)\quad\text{in } D\left[1+t_0,T\right].
$$

Part *(iv)* is trivial, since

$$
\frac{\sqrt{k}}{\log\left(k/(np)\right)}\log\left(\frac{\widehat{x}_p(0,1)}{U_{\mathrm{pre}}(1/p)}\right)=\mathcal{O}_{\mathrm{P}}(1)
$$

and

$$
\frac{\sqrt{k}}{\log\left(k/(np)\right)}\log\left(\frac{\widehat{x}_p(t^*,t^*+t_0)}{U_{\mathrm{post}}(1/p)}\right)=\mathcal{O}_{\mathrm{P}}(1).
$$

The result follows using a similar expansion as in (4.26). □

5 Structural break tests for extremal dependence in β-mixing random vectors

In this chapter we derive a structural break test for the tail event correlation introduced by Hill (2011b). We do so allowing for β-mixing observations, while extant tests (Bücher *et al.*, 2015) require independent data. This allows to test for changes in bivariate extremal dependence *and* serial extremal dependence. As a side product we obtain some central limit theory for the PA-extremogram of Davis *et al.* (2012) that does not require a bootstrap procedure. A simulation study investigates empirical size and power of the test. Finally, we apply our test to the bivariate time series of S&P 500 and DAX log-losses. We find evidence for one break at the beginning of the financial crisis 2007-08, leading to more extremal co-movement.

5.1 Motivation

The study of extremal properties of time series has received considerable attention recently, not least because of the financial crisis of 2007-2008. In financial risk management, for instance, one may be interested in the value-at-risk (VaR), i.e., a small quantile of the return distribution of risky assets. In (re-)insurance interest frequently centers on the tail index as a measure of expected shortfall of log-transformed data. However, these parameters may change over time. Candelon and Stratmans (2006) document tail index changes in returns of emerging market currencies, while Quintos *et al.* (2001) find evidence for such changes in the returns of stock market indices in Malaysia and other Southeast Asian countries during the Asian crisis of 1997-1998.

A less heavily explored strand of research concerns (bivariate) extremal dependence between the components of $\{(X_i, Y_i)'\}_{i\in\mathbb{N}}$. There are many ways to measure extremal dependence. In the case where $Y_i := X_{i-h}$, $h \geq 1$, i.e., serial dependence, the literature suggests, to name a few, the *extremal index* introduced in Leadbetter (1983) as a measure for extremal clustering, the power-law bivariate tail decay model of Ledford and Tawn (2003), and the *extremogram* proposed by Davis and Mikosch (2009). We refer to Hill (2011b, Sec. 6) for definitions of these measures and some discussion. In a more general setting where Y_i is not necessarily a lagged X_i, copula-based notions of (extremal) dependence like the *tail dependence coefficient* have become popular.

While break detection in measures of dependence over the whole real line (e.g., the covariance) has been studied for some time now (e.g., Aue *et al.*, 2009; Wied *et al.*, 2012), only very recently did Bücher *et al.* (2015) propose structural break tests for the tail dependence coefficient. We are not aware of any other tests for breaks in extremal dependence. Yet, the test in Bücher *et al.* (2015) requires data to be independent, which is hardly credible in typical financial applications. Their workaround is to consider filtered residuals. This however has several drawbacks. First, it invalidates their test when dependence changes occur due to parameter shifts in a parametric model instead of changing residual dependence. Second, it is not always clear that a change in the innovation dependence translates into a change in the dependence of the actual observed data, in which one is typically interested in (see Example 5.3 below). Third, one may not always be able to fit a parametric model to obtain estimated residuals (see Section 5.3 below). Fourth, even if one does, there always remains some model risk. A slight misspecification of the model may lead to invalid results, particularly in an extreme value setting as theirs (and ours). For instance, in the context of extreme quantile estimation, Drees (2008, Sec. 2.3) gives an example of a nonlinear 'AR(1)'-model, which is hardly distinguishable from an AR(1) by visual inspection and standard statistical tests. In simulations, he finds that a residual-based quantile estimator yields starkly inferior results to an estimator based on the raw data. Fifth, investigating breaks in serial dependence is not possible because innovations driving the time series are typically i.i.d., whence changes in the serial dependence structure of a times series will generally not be due to changes in the serial dependence structure of the innovations, see, e.g., model (GARCH) below in Section 5.3.

It is, therefore, the purpose of the present chapter to derive structural break tests for the so called *tail event correlation* suggested in Hill (2011b). For the connection to the other measures mentioned above we refer to the discussion in his Section 6.1 and also our Section 5.2.1 below. Our test statistic is almost identical in spirit to that of Bücher *et al.* (2015). We intend to address the above mentioned problems of the test in Bücher *et al.* (2015), stemming from their focus on i.i.d. data, by allowing for β-mixing data. We mention in passing that while our focus is on financial time series, our results can in principle be applied in a wide range of disciplines. We are particularly interested in detecting breaks in the co-movement of asset prices during times of stress, see also Section 5.4. Typically, one expects asset prices to move in lockstep to some degree during a crisis, a phenomenon known as 'diversification meltdown' (Campbell *et al.*, 2008).

The outline of the chapter is as follows. The main results under the null and local alternatives are stated in Section 5.2. Simulation evidence is presented in Section 5.3 and applications in Section 5.4, while proofs are relegated to Section 5.5.

5.2 Main results

This section is organized as follows: Subsection 5.2.1 introduces basic notation and the main assumptions that will be used throughout. Subsection 5.2.2 states convergence results under the null. Results under smooth local alternatives are stated in Subsection 5.2.3. Finally, Section 5.2.4 gives two time series examples satisfying our main assumptions.

5.2.1 Preliminaries

Consider a bivariate $\mathbb{R}^2$-valued stochastic process $\{\boldsymbol{V}_i := (X_i, Y_i)'\}_{i\in\mathbb{N}}$ defined on some probability space $(\Omega, \mathcal{A}, P)$. Let $Z_i \in \{X_i, Y_i\}$ and denote by F_z the time-invariant d.f. of Z_i. We assume $1 - F_z$ to be regularly varying with parameter $-\alpha_z < 0$ (written $1 - F_z \in RV_{-\alpha_z}$), i.e.,

$$1 - F_z(y) = y^{-\alpha_z} L_z(y), \qquad y > 0. \tag{5.1}$$

Here, α_z is called the tail index of Z_i and $L_z(\cdot)$ is slowly varying, i.e.,

$$\frac{L_z(\lambda y)}{L_z(y)} \underset{(y\to\infty)}{\longrightarrow} 1 \qquad \forall\, \lambda > 0.$$

Remark 5.1. (a) The types of power laws in (5.1) for the tails of random variables are very common in economics and finance, ranging from income to stock market returns to trading volume and beyond (see Gabaix, 2009, for an extensive overview). They are also encountered in internet-traffic and insurance claim data (see Resnick, 2007).

(b) Not only are regularly varying d.f.s an often credible model for the tails of empirical data, but they also occur naturally in theory in the study of maxima and sums of i.i.d. random variables and as the tails of solutions to stochastic recurrence equations (see Mikosch, 2005, Sec. 1).

(c) Even under the alternative of a change in the extremal dependence we maintain the assumption of stationary marginals. Bücher *et al.* (2015) attempted to weaken that assumption in their independent setting, yet limit distributions turned out to be intractable and the test statistic depends on unknown parameters. We leave this task in our dependent setting for future study.

Define

$$b_z(t) := F_z^{\leftarrow}\left(1 - \frac{1}{t}\right), \qquad t > 1,$$

as the $(1 - 1/t)$-quantile, where $^{\leftarrow}$ denotes the left-continuous inverse. Denote by

$Z_{(1)} \geq \ldots \geq Z_{(n)}$ the order statistics of $Z_1, \ldots, Z_n$.

As we will only be interested in tail dependence, we need some intermediate sequence $k = k_n \in \mathbb{N}$ with $k \leq n-1$, i.e.,

$$k \underset{(n\to\infty)}{\longrightarrow} \infty \qquad \text{and} \qquad \frac{k}{n} \underset{(n\to\infty)}{\longrightarrow} 0,$$

specifying where 'the tail begins'. This sequence controls the number of large observations used in the estimation of the extremal dependence of X_i and Y_i.

As mentioned in the motivation, Hill (2011b, Sec. 5.2) recommends the *tail event correlation* as a measure of tail dependence, which is defined as

$$r_{n,i} := \frac{n}{k}\left[\mathrm{P}\left\{X_i > b_x(n/k), Y_i > b_y(n/k)\right\} - \mathrm{P}\left\{X_i > b_x(n/k)\right\}\mathrm{P}\left\{Y_i > b_y(n/k)\right\}\right].$$

A natural estimator of $r_{n,i}$ is given by

$$\widehat{r}_n := \frac{1}{k}\sum_{i=1}^{n}\left[I_{\left\{X_i > X_{(k+1)}, Y_i > Y_{(k+1)}\right\}} - \left(\frac{k}{n}\right)^2\right]. \tag{5.2}$$

Note that under (5.1) it holds that $\mathrm{P}\left\{Z_i > b_z(n/k)\right\} \sim k/n$. We now briefly discuss the connections to some of the other dependence measures mentioned in the Introduction.

Remark 5.2. (a) If $Y_i = X_{i-h}$ for a stationary sequence of r.v.s $\{X_i\}$ and the limit of $r_{n,i}$ exists as $n \to \infty$, then this limit is called *extremogram* in Davis and Mikosch (2009, Sec. 1.4). Notice that the estimator $\widehat{r}_n$, unlike the estimator of Davis and Mikosch (2009), has the advantage of using a data dependent threshold.

(b) If the d.f.s of the marginals F_x and F_y are continuous, then by a standard probability integral transformation

$$\begin{aligned} r_{n,i} &\sim \frac{n}{k}\mathrm{P}\left\{F_x(X_i) > 1-k/n, F_y(Y_i) > 1-k/n\right\} \\ &= \mathrm{P}\left\{F_x(X_i) > 1-k/n \mid F_y(Y_i) > 1-k/n\right\}, \end{aligned}$$

where $a_n \sim b_n$ is taken to mean $\lim_{n\to\infty} a_n/b_n = 1$. The limit of this latter probability (if it exists) is called upper tail dependence coefficient (TDC), see also Frahm *et al.* (2005, Sec. 2). It is exactly this TDC that Bücher *et al.* (2015) investigate. We however focus on its 'pre-asymptotic' version $r_{n,i}$ because of two reasons. First, as pointed out in Hill (2011b, Sec. 6.2), it better captures what can be construed as dependence. For instance, if X_i are independent, then $r_{n,i} = 0$ for $Y_i = X_{i-h}$, whence the TDC is also equal to zero. Yet, if X_i follows the Gaussian log-autoregressive stochastic volatility process of Hill (2011b, Sec. 6.2), then the TDC is also zero, while clearly some extremal serial dependence is present. In fact, $\lim_{n\to\infty}\frac{n}{k}r_{n,i} > 0$. Hence, the TDC does not capture the different dependence structures, while its pre-asymptotic version does. The second reason for our focus on $r_{n,i}$ is that it is a conditional probability, 'which is often the quantity of primary interest in applications' (Davis *et al.*, 2012, p. 143).

If it is in doubt whether extremal dependence in $(X_1, Y_1)', \ldots, (X_n, Y_n)'$ is constant over time, one may wish to test the following hypothesis:

$$\begin{aligned} \mathcal{H}_0: &\quad r_{n,1} = \ldots = r_{n,n} \qquad \text{versus} \\ \mathcal{H}_1: &\quad \text{Not } \mathcal{H}_0. \end{aligned}$$

The dependence concept used here is that of β-mixing. Recall that a (possibly triangular) sequence of random vectors $\{\boldsymbol{V}_{n,i}\}_{n\in\mathbb{N}, i=1,\ldots,n}$ is β-mixing if

$$\beta_n(l) := \sup_{m\in\{1,\ldots,n-l-1\}} \mathbb{E}\left[\sup_{A\in\mathcal{F}^n_{n,m+l+1}} \left|\mathrm{P}\{A|\mathcal{F}^m_{n,1}\} - \mathrm{P}\{A\}\right|\right] \underset{(l\to\infty)}{\longrightarrow} 0, \tag{5.3}$$

where $\mathcal{F}^m_{n,l}$ is the σ-algebra generated by $\{\boldsymbol{V}_{n,l}, \ldots, \boldsymbol{V}_{n,m}\}$. We allow for a triangular array to also cover the alternative in our development. Of course, if $\boldsymbol{V}_{n,i} = \boldsymbol{V}_i$ is in fact a *sequence* of β-mixing random vectors, then array β-mixing in the sense of (5.3) follows.

We now state our main assumptions that will be maintained throughout.

(D1) $\{\boldsymbol{V}_{n,i}\}_{i\in\mathbb{N}}$ is a β-mixing process with mixing coefficients $\beta_n(\cdot)$, such that

$$\lim_{n\to\infty} \frac{n}{r_n}\beta_n(l_n) + \frac{r_n}{\sqrt{k}} + \frac{r_n k}{n} = 0$$

for sequences $\{l_n\}_{n\in\mathbb{N}} \subset \mathbb{N}, \{r_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$ tending to infinity with $l_n^2 = o(r_n)$ and $r_n = o(n)$. Furthermore, $\boldsymbol{V}_{n,i}$ has stationary marginal distributions.

(D2) It holds uniformly in $j \geq 0$ that

$$\lim_{n\to\infty} \frac{n}{r_n k} \operatorname{Var}\left(\sum_{i=j+1}^{j+r_n} I_{\left\{X_i > b_x\left(\frac{n}{k}\right), Y_i > b_y\left(\frac{n}{k}\right)\right\}}\right) = \sigma^2 > 0. \tag{5.4}$$

(D3a) For the slowly varying functions $L_z(\cdot)$ from (5.1) there exist (possibly different) positive measurable functions $g(\cdot)$ with bounded increase, i.e.,

$$\frac{g(\lambda y)}{g(y)} \leq C\lambda^{\tau} \quad \text{for } \lambda \geq 1, y \geq y_0, \quad \text{where } C, y_0 < \infty, \tau \leq 0,$$

such that

$$\frac{L(\lambda y)}{L(y)} - 1 \underset{(y \to \infty)}{=} \mathcal{O}(g(y)), \qquad \forall\, \lambda > 0, \tag{5.5}$$

where $\sqrt{k} g(b_z(n/k)) \to 0$ as $n \to \infty$.

(D3b) For the slowly varying functions $L_z(\cdot)$ from (5.1) there exist (possibly different) $\kappa(y) = K \int_1^y t^{\gamma-1} \mathrm{d}t$, $K \in \mathbb{R}$, and a positive measurable function $g(\cdot) \in RV_\gamma$, $\gamma \leq 0$, such that

$$\lim_{y\to\infty} \frac{\frac{L(\lambda y)}{L(y)} - 1}{g(y)} = \kappa(\lambda) \qquad \forall\, \lambda > 0, \tag{5.6}$$

where $\sqrt{k} g(b_z(n/k)) \to A \in \mathbb{R}$ as $n \to \infty$.

Remark 5.3. (a) Condition **(D1)** allows for the application of a standard 'big block - small block' argument similarly as in Chapter 2, where the small blocks of length l_n are asymptotically negligible and the big blocks of length r_n converge to some well-defined limit by virtue of **(D2)**. We refer to Proposition 5.1 below for some more easily verified sufficient conditions for **(D2)**.

(b) **(D3)** (which henceforth is taken to mean either **(D3a)** or **(D3b)**) is a widely-used second-order condition (e.g., in Hsing, 1991; Hill, 2011b) that controls the speed of convergence in (5.1). Some examples of distribution functions satisfying **(D3)** are given in Haeusler and Teugels (1985). For instance, in their Example 1 they show that a d.f. satisfying

$$1 - F(x) = Cx^{-\alpha}(1 + \mathcal{O}(x^{-\beta})) \quad \text{as} \quad x \to \infty, \quad C, \alpha, \beta > 0$$

(which is only slightly stronger than (5.1)) satisfies **(D3a)** with $k = o\left(n^{\frac{2\beta}{2\beta+\alpha}}\right)$.

(c) Note that we do not require a finite moment assumption. This contrasts with change point tests based on sample autocovariances, which require finite 4th moments for their central limit theory (Aue *et al.*, 2009; Wied *et al.*, 2012). This may not always be satisfied in financial applications where typically $\alpha \in (2, 4)$, implying infinite 4th moments (Resnick, 2007, Sec. 9.2.1). Such an assumption is even less credible for insurance data, where variances may not exist with $\alpha \in (1, 2)$ (Resnick, 2007, Figs. 4.7 & 4.16).

Conditions **(D1)-(D3)** relax some assumptions of the test proposed in Bücher *et al.* (2015). For instance, our scheme covers a wide range of short-memory processes. Allowing for dependence is essential as, under dependence, the limit distribution of the test statistic considered in Bücher *et al.* (2015) will be scaled by some (dependence-structure dependent) factor. They perform their tests for dependent data on pre-filtered residuals, arguing that they are almost i.i.d. This may be questionable in several respects, as detailed in the motivation. Furthermore, unlike Bücher *et al.* (2015, Assumption 3.2(b)), we do not require a speed of convergence in (5.10) below.

Proposition 5.1. *Suppose that $\boldsymbol{V}_{n,i} = \boldsymbol{V}_i$ is a strictly stationary sequence of random vectors and that for all $m \in \mathbb{N}_0$*

$$\frac{n}{k}\mathrm{P}\left\{X_1 > b_x(n/k), Y_1 > b_y(n/k), X_{1+m} > b_x(n/k), Y_{1+m} > b_y(n/k)\right\} \underset{(n\to\infty)}{\longrightarrow} c_m, \tag{5.7}$$

where $c_0 > 0$, and

$$\lim_{h\to\infty} \limsup_{n\to\infty} \frac{n}{k} \sum_{m=h+1}^{r_n} \mathrm{P}\{X_1 > b_x(n/k), Y_1 > b_y(n/k), X_{1+m} > b_x(n/k), Y_{1+m} > b_y(n/k)\} = 0. \tag{5.8}$$

Then **(C2)** *is met with $\sigma^2 = c_0 + 2\sum_{m=1}^{\infty} c_m > 0$ if $\lim_{n\to\infty} r_n k/n = 0$.*

Remark 5.4. (a) Note that the assumption $c_0 > 0$ in (5.7) prohibits asymptotic independence of X_i and Y_i, in the sense that

$$\lim_{n\to\infty} \mathrm{P}\left\{X_i > b_x(n/k) \mid Y_i > b_y(n/k)\right\} = 0 \tag{5.9}$$

as $\mathrm{P}\left\{Y_i > b_y(n/k)\right\} \sim k/n$. The equivalent of this condition in Bücher *et al.* (2015) is the requirement that $\Delta_i \neq 0$ in their Assumption 3.1. Note further that condition (5.7) for $m = 0$ reads as

$$\frac{n}{k}\mathrm{P}\left\{X_1 > b_x(n/k), Y_1 > b_y(n/k)\right\} \underset{(n\to\infty)}{\longrightarrow} c_0 > 0, \tag{5.10}$$

i.e., that the left-hand side of this condition - exactly the quantity in which we aim to detect changes - has a well-defined limit.

(b) If $\boldsymbol{V}_i = (X_i, X_{i-h})'$, where $(X_i)_{i\in\mathbb{Z}}$ is strictly stationary and multivariate regular varying, then (5.7) is trivially satisfied. See Fasen *et al.* (2010, p. 3 and p. 7), where the limits c_m are values of the extreme dependence function introduced in their Definition 1.1. The advantage is that multivariate regular variation has been verified for a large array of time series, e.g., MA(∞) processes with heavy-tailed noise, GARCH(1,1) processes, AR(1) processes with ARCH(1) errors; see Fasen *et al.* (2010, Thms. 3.3, 3.9, 3.14) and the preceding lemmas.

(c) An almost trivial, yet quite useful, sufficient condition for (5.8) is

$$\lim_{h\to\infty} \limsup_{n\to\infty} \frac{n}{k} \sum_{m=h+1}^{r_n} \mathrm{P}\left\{X_1 > b_x(n/k), X_{1+m} > b_x(n/k)\right\} = 0. \tag{5.11}$$

5.2.2 Results under the null

Our test is based on comparing different subsample estimates of $r_{n,i}$ à la

$$G_n(t) := \sqrt{k}t(1-t)\left\{\frac{1}{kt}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>X_{(k+1)},Y_i>Y_{(k+1)}\right\}} - \frac{1}{k(1-t)}\sum_{i=\lfloor nt\rfloor+1}^{n} I_{\left\{X_i>X_{(k+1)},Y_i>Y_{(k+1)}\right\}}\right\}. \tag{5.12}$$

Note that the (constant) marginal tail probabilities cancel out here.

Theorem 5.1. *Suppose assumptions* **(D1)**-**(D3)** *are met. Then under* $\mathcal{H}_0$

$$G_n(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma B(t) \qquad \text{in } D[0,1],$$

where $\{B(t)\}_{t\in[0,1]}$ *denotes a standard Brownian bridge.*

Here and in the following $D[0,1]$ denotes the space of $\mathbb{R}$-valued càdlàg functions equipped with the Skorohod metric and the Borel σ-field $\mathcal{D}[0,1]$. From the above theorem convergence results follow easily for a wealth of test statistics by virtue of the continuous mapping theorem (CMT). For our comparative simulation study in Section 5.3 we will focus on the Crámer-von Mises functional

$$\mathcal{T}_n := \widehat{\sigma}^{-2}\int_0^1 G_n(t)^2 \mathrm{d}t \xrightarrow[(n\to\infty)]{\mathcal{D}} \int_0^1 B(t)^2 \mathrm{d}t \tag{5.13}$$

of $G_n(t)$, Bücher *et al.* (2015, Eq. (8)) report good finite-sample properties for this functional. Quantiles of $\int_0^1 B(t)^2 \mathrm{d}t$, which serve as critical values for our test, were tabulated in Anderson and Darling (1952, Table 1) and are repeated in Table 5.1. Here, $\widehat{\sigma}^2$ denotes a consistent estimator of σ^2. Relation (5.4) suggests the following estimator $\widetilde{\sigma}^2$ of σ^2

$$\sum_{j=1}^{\lfloor n/r_n \rfloor} \left[\sum_{i=(j-1)r_n+1}^{jr_n} \frac{1}{\sqrt{k}} I_{\{X_i > b_x(n/k), Y_i > b_y(n/k)\}} - \frac{r_n}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{k}} I_{\{X_i > b_x(n/k), Y_i > b_y(n/k)\}} \right]^2 \tag{5.14}$$

similarly as in Davis and Mikosch (2009, p. 1006). Again, we have to show that $b_x(n/k)$ and $b_y(n/k)$ can be replaced by $X_{(k+1)}$ and $Y_{(k+1)}$ for the estimator to be feasible. This gives the estimator defined in the following

Theorem 5.2. *Suppose assumptions* **(D1)**-**(D3)** *are met. Then under $\mathcal{H}_0$ and, if additionally $n/k^{3/2} \to 0$ holds, also under $\mathcal{H}_1$*

$$\widehat{\sigma}^2 \xrightarrow[(n\to\infty)]{\mathrm{P}} \sigma^2,$$

where $\widehat{\sigma}^2$ is defined by

$$\sum_{j=1}^{\lfloor n/r_n \rfloor} \left[\sum_{i=(j-1)r_n+1}^{jr_n} \frac{1}{\sqrt{k}} I_{\{X_i > X_{(k+1)}, Y_i > Y_{(k+1)}\}} - \frac{r_n}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{k}} I_{\{X_i > X_{(k+1)}, Y_i > Y_{(k+1)}\}} \right]^2 .$$

Remark 5.5. The condition $n/k^{3/2} \to 0$ is mild: even under the null of no change it was imposed in Davis and Mikosch (2009, Thm. 3.2) and Hill (2011b, Sec. 5.2), who basically studied the same estimator of $\widehat{r}_n$.

As a side result we obtain the following

Corollary 5.1. *Suppose assumptions* **(D1)**-**(D3)** *are met for $\mathbf{V}_i = (X_i, X_{i-h})'$. Then under $\mathcal{H}_0$*

$$\frac{\sqrt{k}}{\widehat{\sigma}} \left\{ \frac{1}{k} \sum_{i=1}^{n} I_{\{X_i > b_x(n/k), X_{i-h} > b_x(n/k)\}} - \frac{n}{k} \mathrm{P}\left\{ X_i > b_x(n/k), X_{i-h} > b_x(n/k) \right\} \right\} \xrightarrow[(n\to\infty)]{\mathcal{D}} \mathcal{N}(0,1). \tag{5.15}$$

α_q	0.50	0.60	0.70	0.80	0.90	0.95	0.99
$c(\alpha_q)$	0.11888	0.14663	0.18433	0.24124	0.34730	0.46136	0.74346

Table 5.1: $c(\alpha_q) = \alpha_q$-quantile of $\int_0^1 B(t)^2 \mathrm{d}t$

The quantities in brackets of the left-hand side of (5.15) are the sample extremogram (left) and the pre-asymptotic extremogram (PA-extremogram, right) introduced in Davis *et al.* (2012, Eqs. (1.2) and (1.4)) for the particular choices $A = B = (1, \infty)$ and $a_m = b_x(n/k)$. By premultiplying the X_i with -1, Corollary 5.1 also provides central limit theory for the choices $A = B = (-\infty, -1)$. Thus, it covers the arguably most popular choices of A and B for the PA-extremogram (see Davis *et al.*, 2012, Sec. 3). Unlike in Davis *et al.* (2012), using Corollary 5.1 one does not require a bootstrap procedure to construct confidence intervals.

Furthermore, by definition of our test statistic in (5.13) our test may also be used as a test for constancy of the PA-extremogram. Thus, the following integrated procedure for estimation and inference for the PA-extremogram suggests itself. First, test for a change in the serial extremal dependence structure. If no break is detected, then use Corollary 5.1 for estimation and inference. See Section 5.4.2 for an application.

5.2.3 Results under the alternative

We will explore the behavior of our test under local alternatives:

$$\mathcal{H}_1: \quad r_{n,i} = r_{n,0} + \frac{1}{\sqrt{k}} g_n\left(\frac{i}{n}\right), \qquad i = 1, \ldots, n. \tag{5.16}$$

The form of the local alternative is inspired by Wied *et al.* (2012, Sec. 3). The class of functions $\{g_n(\cdot)\}_{n\in\mathbb{N}}$ must belong to $\mathcal{G}$, which is defined as the set of uniformly bounded sequences of functions $\{g_n(\cdot) : [0,1] \to \mathbb{R}\}$ satisfying

$$\lim_{n\to\infty} \sup_{t\in[0,1]} \left| \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} g_n(i/n) - \int_0^t g(z)\mathrm{d}z \right| = 0$$

for some integrable $g(\cdot)$. For instance, if all $g_n(\cdot)$ are Riemann-integrable and the $g_n(\cdot)$ converge uniformly to some function $g(\cdot)$, then $g(\cdot)$ will be Riemann-integrable as well and $\{g_n\} \in \mathcal{G}$. The assumption $\{g_n\} \in \mathcal{G}$ ensures a well-defined limit in (5.41) below.

Theorem 5.3. *Suppose assumptions* **(D1)**-**(D3)** *are met. Then under* $\mathcal{H}_1$ *for* $\{g_n\} \in \mathcal{G}$ *such that*

$$C(t) := \int_0^t g(z)\mathrm{d}z - t\int_0^1 g(z)\mathrm{d}z \neq 0 \qquad \textit{for some } t \in [0,1],$$

we have

$$G_n(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma B(t) + C(t) \qquad \textit{in } D[0,1],$$

so in particular

$$\mathcal{T}_n \xrightarrow[(n\to\infty)]{\mathcal{D}} \int_0^1 \left(B(t) + C(t)/\sigma\right)^2 \mathrm{d}t. \tag{5.17}$$

The above theorem is similar in spirit to Wied *et al.* (2012, Thm. 2). By arguments as in Rothe and Wied (2013, Proof of Theorem 3 (iii)) one can deduce from (5.17) that the asymptotic level of our test is never smaller than α_q under local alternatives (part *(i)* of the next corollary). Further, let $g_n(\cdot) = Mh_n(\cdot)$, where h_n captures the functional form of the alternative and $M > 0$ its magnitude. Then we may argue as in Wied *et al.* (2012, Cor. 1) to derive part *(ii)* of

Corollary 5.2. *Suppose the assumptions of Theorem 5.3 are met. If $g_n(\cdot) = Mh_n(\cdot)$, then*

(i)

$$\lim_{n\to\infty} \mathrm{P}\{\mathcal{T}_n > c(\alpha_q)\} \geq \alpha_q;$$

(ii) for any $\varepsilon > 0$ there exists $M_0 = M_0(\varepsilon) > 0$ such that

$$\lim_{n\to\infty} \mathrm{P}\{\mathcal{T}_n > c(\alpha_q)\} > 1 - \varepsilon \qquad \forall\, M > M_0.$$

The above corollary shows that our test has local power and, as the magnitude of the changes M increases, so does the power of our test.

5.2.4 Two examples

In this section we discuss specific models that satisfy **(D1)-(D2)**. Example 5.1 deals with the special case $Y_i = X_{i-h}$, i.e., serial dependence, and Example 5.2 covers dependence in a multivariate stochastic volatility (SV) model. While the first-order condition in (5.1) is satisfied for both examples, the second-order condition **(D3)** has, to the best of our knowledge, not yet been universally verified.

Example 5.1 (GARCH)**.** Consider a GARCH(1,1) process $\{X_i\}_{i\in\mathbb{N}_0}$ defined by the multiplicative model

$$X_i = \sigma_i Z_i, \qquad i \in \mathbb{N}_0,$$

where $\{Z_i\}_{i\in\mathbb{N}_0}$ is an i.i.d. sequence independent of σ_0 and $\left\{\sigma_i^2\right\}_{i\in\mathbb{N}_0}$ is the solution of

$$\sigma_i^2 = \alpha_0 + \alpha_1 X_{i-1}^2 + \beta\sigma_{i-1}^2, \qquad i \in \mathbb{N},$$

with parameter restrictions $\alpha_0, \alpha_1 > 0$ and $\beta \in [0, 1)$. The case $\beta = 0$ corresponds to an ARCH(1) process. Now, suppose that Z_0 has a positive density on $\mathbb{R}$, and either $\mathbb{E}|Z_0|^h < \infty$ for $h \in (0, h_0)$ and $\mathbb{E}|Z_0|^h = \infty$ for $h \geq h_0 > 0$, or $\mathbb{E}|Z_0|^h < \infty$ for all

$h > 0$. Additionally assume that

$$\mathbb{E}[\log(\alpha_1 Z_0^2 + \beta)] < 0.$$

Under these assumptions geometric mixing of $\{|X_i|\}$ (and hence that of $\boldsymbol{V}_i = (|X_i|, |X_{i-h}|)'$) follows from Lindner (2009, Thm. 8), i.e., $\beta(n) \le C\eta^n$ for some $C > 0, \eta \in (0,1)$. Furthermore, the tail index $\alpha > 0$ of (the strictly stationary) $|X_i|$ is the unique solution to

$$\mathbb{E}[\alpha_1 Z_0^2 + \beta]^{\alpha/2} = 1,$$

see Fasen *et al.* (2010, Lem. 3.8). However, much less is known about the second-order properties **(D3)** of $|X_i|$, although some progress has been made in Baek *et al.* (2009). For an ARCH(1) process we have (see Drees, 2003, p. 634)

$$\mathrm{P}\{|X_i| > x\} = \mathrm{P}\{X_i^2 > x^2\} = dx^{-2\alpha}\left(1 + \mathcal{O}(x^{-2\alpha\tau})\right) \qquad \text{for some } d, \tau > 0. \quad (5.18)$$

Then by Haeusler and Teugels (1985, Example 1) assumption **(D3a)** is fulfilled if $k = o(n^{2\tau/(2\tau+1)})$.

Now if (5.18) also holds for the GARCH(1,1) process as has been frequently assumed in the literature (cf. Sun and Zhou, 2014, Rem. 10), then conditions **(D1)-(D3)** are satisfied for $\boldsymbol{V}_i = (|X_i|, |X_{i-h}|)'$, $h \in \mathbb{N}$, for the following choices:

$$k = n^{1-\delta} \text{ for } \delta \in \left(1 - \frac{2\tau}{2\tau+1}, 1\right), \quad r_n = n^{\nu} \text{ for } \nu < \delta/2, \quad l_n = -2\frac{\log n}{\log \eta}.$$

Thus, **(D1)** can be easily seen to be satisfied by geometric mixing. As for **(D2)**, we note that (5.11) is satisfied by analogous calculations as in Davis and Mikosch (2009, Sec. 4.1) and (5.7) due to Fasen *et al.* (2010, Thm. 3.9). More specifically, for $m = 0$ (5.7) reads as

$$\frac{n}{k}\mathrm{P}\left\{|X_h| > b_{|x|}(n/k), |X_0| > b_{|x|}(n/k)\right\} \underset{(n\to\infty)}{\longrightarrow} \frac{\mathbb{E}\left[\min\left(|Z_0|^{\alpha}, |Z_h|^{\alpha}\prod_{i=1}^{h}(\alpha_1 Z_{i-1}^2 + \beta)^{\alpha/2}\right)\right]}{\mathbb{E}|Z_0|^{\alpha}}. \quad (5.19)$$

This suggests that the extremal dependence is influenced by the distribution of Z_0 and the parameters α_1 and β, but not α_0.

Example 5.2 (Multivariate SV)**.** We take up Example 2.4 from Janssen and Drees (2016). Consider their bivariate stochastic volatility model $\{\boldsymbol{V}_i = (X_i, Y_i)'\}_{i\in\mathbb{N}}$ and adapt it slightly to exhibit positive dependence between the components so as to be consistent with our exposition so far. Concretely,

$$\boldsymbol{V}_i = \boldsymbol{A}_i \boldsymbol{Z}_i,$$

where

$$\boldsymbol{A}_i := \begin{pmatrix} \exp(h_{1,i}/2) & 0 \\ \exp((h_{1,i}+q_i)/2) & \exp(h_{2,i}/2) \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{Z}_i := \begin{pmatrix} \epsilon_{1,i} \\ \epsilon_{2,i} \end{pmatrix},$$

and $q_i, h_{1,i}, h_{2,i}$ follow stationary AR(1)-models

$$\begin{aligned} q_{i+1} &= \alpha_q + \beta_q q_i + u_i, \qquad i \in \mathbb{Z}, \\ h_{m,i+1} &= \mu_m + \phi_m h_{m,i} + \eta_{m,i}, \qquad i \in \mathbb{Z},\ m = 1, 2, \end{aligned}$$

with $\beta_q, \phi_m \in (-1, 1)$, $\eta_{m,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_{\eta,m})$, $u_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_u)$ all mutually independent. Further, assume that $\boldsymbol{Z}_i$ is i.i.d. and multivariate regular varying with limit measure $\mu\{\cdot\}$ on $[-\infty, \infty]^2 \setminus \{\boldsymbol{0}\}$, i.e.,

$$t\mathrm{P}\{b(t)^{-1}\boldsymbol{Z} \in \cdot\} \underset{(n\to\infty)}{\overset{v}{\longrightarrow}} \mu\{\cdot\}, \qquad \text{for some function } b(t) \to \infty, \tag{5.20}$$

where $\overset{v}{\longrightarrow}$ denotes vague convergence and $\mu\{\cdot\}$ is a Radon measure that is not identically zero and that is not degenerate at a point, see, e.g., Resnick (2007, Ch. 6, in particular Thm. 6.1). Recall that $\mu\{\cdot\}$ is necessarily homogenous, $\mu\{t\boldsymbol{B}\} = t^{-\alpha}\mu\{\boldsymbol{B}\}$ for $t > 0$ and $\boldsymbol{B}$ a Borel-set of $[-\infty, \infty]^2 \setminus \{\boldsymbol{0}\}$, where $\alpha > 0$ is called the *index* of regular variation. We assume that $\mu\{\cdot\}$ puts positive mass on and off the axes of the non-negative orthant, such that the components of $\boldsymbol{Z}_i$ are not asymptotically independent, and all $\boldsymbol{x} = (x_1, x_2)$ with $x_1 > 0$ are continuity points of the function $\boldsymbol{x} \mapsto \mu\{(x_1, \infty] \times (x_2, \infty]\}$. These assumptions will ensure that (5.7) holds.

Note that $\{q_i\}$ and $\{h_{m,i}\}$ are strictly stationary AR(1)-models with stationary distributions

$$Q \sim \mathcal{N}\left(\frac{\alpha_q}{1-\beta_q}, \frac{\sigma^2_u}{1-\beta^2_q}\right) \qquad \text{and} \qquad H_m \sim \mathcal{N}\left(\frac{\mu_m}{1-\phi_m}, \frac{\sigma^2_{\eta,m}}{1-\phi^2_m}\right), \quad m = 1, 2,$$

whence $\{\boldsymbol{V}_i\}$ is also strictly stationary. Furthermore, $\{q_i\}$ and $\{h_{m,i}\}$ are geometrically β-mixing (cf., e.g., Bradley, 1986, Ex. 6.1). Taking exp of these processes still leaves us with geometrically β-mixing time series (e.g., Bradley, 1986, p. 170). Using the independence of the q_i and $h_{m,i}$, this carries over to $\boldsymbol{V}_i$ by standard mixing results (cf., e.g., Bradley, 1986, Sec. 3).

For the verification of **(D1)**-**(D3)** write

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} \exp(h_{1,i}/2)\epsilon_{1,i} \\ \exp(h_{2,i}/2)\epsilon_{2,i} + \exp((h_{1,i}+q_i)/2)\epsilon_{1,i} \end{pmatrix}$$

and define

$$\boldsymbol{A}^{-1} = \begin{pmatrix} \exp(-H_1/2) & 0 \\ -\exp((Q-H_2)/2) & \exp(-H_2/2) \end{pmatrix}.$$

By Basrak *et al.* (2002, Prop. A.1) the vector $(X_i, Y_i)'$ is multivariate regularly varying with index α and limit measure $\widetilde{\mu}\{\cdot\} := \mathbb{E}[\mu\{\boldsymbol{A}^{-1}(\cdot)\}]$. Thus, the components X_i and Y_i satisfy (5.1) with tail index α. Hence, the heavy tails are inherited from the $\epsilon_{m,i}$'s, whose tail indices of course also equal α, as well as their extremal dependence structure as represented by $\mu\{\cdot\}$. For the verification of **(D3)** assume additionally that

$$1 - F_{\epsilon_m}(x) = d_m x^{-\alpha} \exp\left(\int_1^x \frac{\eta_m(s)}{s} \mathrm{d}s\right), \qquad x > 0,\ c > 0$$

for some $\eta_m(s) = \mathcal{O}(s^{\alpha\rho_m})$, $\rho_m < 0$ (covering Fréchet-, t_ν- and generalized Pareto distributions). Then, by arguments in Section 4.2.3,

$$\mathrm{P}\{X_i > x\} = d_X x^{-\alpha}\left(1 + \mathcal{O}(x^{-\beta_X})\right) \qquad \text{for some } d_X, \beta_X > 0, \tag{5.21}$$

whence **(D3a)** is satisfied. Of course, the survivor functions of the two (independent) summands defining Y_i also satisfy such an expansion. However, to the best of our knowledge, it is not known if **(D3a)** is preserved under convolution. As results in Geluk *et al.* (2000) suggest that this may be the case, we assume a similar expansions as in (5.21) also holds for Y_i (with $d_Y, \beta_Y > 0$), such that, by Haeusler and Teugels (1985, Example 1) again, **(D3a)** is satisfied for $k = o(n^{2\beta/(2\beta+\alpha)})$ for $\beta = \min(\beta_X, \beta_Y)$.

For the following choices the remaining conditions **(D1)** and **(D2)** can be shown to be met:

$$k = n^\delta, \quad r_n = n^{\delta/3} \text{ for } \delta < \min\left(\frac{2\beta}{2\beta+\alpha}, \frac{1}{1+1/3}\right), \quad l_n = -2\frac{\log n}{\log \eta}.$$

Assumption **(D1)** again follows from geometric mixing. For condition **(D2)** it suffices to verify, first, that (5.11) holds, where of course the X_i in the first component are generated according to a simple univariate stochastic volatility model. For such a model this condition was checked in Davis and Mikosch (2009, Sec. 4.2). Second, for $m = 0$ condition (5.7) is a consequence of the multivariate regular variation of $\boldsymbol{V}_1$ from Basrak *et al.* (2002, Prop. A.1). In particular, we obtain with $b(\cdot)$ from (5.20) that for $x, y > 0$

$$\frac{n}{k}\mathrm{P}\{X_i > xb_x(n/k), Y_i > yb_x(n/k)\} \underset{(n\to\infty)}{\longrightarrow} \mathbb{E}[\mu\{\boldsymbol{A}^{-1}((x,\infty] \times (y,\infty])\}]$$
$$= \mathbb{E}[\mu\{(\exp(-H_1/2)x,\infty] \times (-\exp((Q-H_2)/2)x + \exp(-H_2/2)y,\infty]\}], \quad (5.22)$$

where the limit is strictly larger than 0, because of the properties of $\mu\{\cdot\}$. Furthermore, because X_i is scaled both by $b(n/k)$ (by multivariate regular variation of V_i and the fact that $\mu\{\cdot\}$ spreads mass on both axes, see also Resnick (2007, p. 175)) and $b_x(n/k)$ (by Resnick (2007, Sec. 2.2.1)), we get convergence of

$$\frac{n}{k}\mathrm{P}\{X_i > xb(n/k)\} \qquad \text{and} \qquad \frac{n}{k}\mathrm{P}\{X_i > xb_x(n/k)\}$$

to non-degenerate limits. By the convergence of types theorem (e.g., Billingsley, 1995, Thm. 14.2) we then get $\lim_{n\to\infty} b_x(n/k)/b(n/k) = a_x > 0$ and by similar arguments $\lim_{n\to\infty} b_y(n/k)/b(n/k) = a_y > 0$. Hence, from (5.22) and properties of $\mu\{\cdot\}$

$$\frac{n}{k}\mathrm{P}\{X_i > b_x(n/k), Y_i > b_y(n/k)\} \underset{(n\to\infty)}{\longrightarrow} \mathbb{E}[\mu\{\boldsymbol{A}^{-1}((a_x,\infty] \times (a_y,\infty])\}] > 0. \quad (5.23)$$

For $m \in \mathbb{N}$ write

$$\begin{pmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_{1+m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{1+m} \end{pmatrix} \begin{pmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_{1+m} \end{pmatrix}.$$

Note that the components of $(\boldsymbol{Z}_1, \boldsymbol{Z}_{1+m})'$ are independent, such that multivariate regular variation is preserved under binding (cf. Resnick, 2007, Lemma 7.2). Then again (5.7) follows from Basrak *et al.* (2002, Prop. A.1).

Remark 5.6. How to check the conditions on $\mu\{\cdot\}$ in Example 5.2 for a given bivariate distribution? For this note that is sufficient for $\mu\{\cdot\}$ to have a positive Lebesgue-density on $\mathbb{R}^2 \setminus \{\boldsymbol{0}\}$. Then Resnick (2007, Secs. 6.5.3 & 6.5.4) can be of help. There it was shown that, e.g., the bivariate t-distribution has such a limiting measure, see also de Haan and Resnick (1987, Cor. 3.4).

5.3 Simulations

This section investigates the finite-sample properties of our test in the context of serial dependence (for Example 5.1) and bivariate dependence (for Example 5.2). Furthermore, we compare our test with the $\mathcal{S}_n$-test proposed by Bücher *et al.* (2015, Cor. 3.4), which is based on estimated residuals if data cannot be credibly assumed to be i.i.d. We also want to investigate how robust our results are with respect to the choice of k, which can be a delicate issue. A further parameter that is only specified asymptotically is r_n. In order to keep the presentation concise we do not compare results for different choices of r_n. Unreported simulations reveal that there seems to be a trade-off between size and power - higher values of r_n lead to better finite-sample size, yet lower power. These suggest that choosing $r_n = \left\lfloor 15 \cdot k^{0.2} \right\rfloor$ always yields a good compromise. This choice was inspired by the growth condition for r_n from the two examples in Section 5.2.4.

Concretely, the data generating process (DGP) is the following GARCH(1,1)-model:

$$X_i = \sigma_i Z_i, \qquad \text{where} \qquad \begin{cases} Z_i & \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \\ \sigma_i^2 & = 0.01 + \alpha_1^{(i)} X_{i-1}^2 + 0.15\sigma_{i-1}^2, \end{cases} \tag{GARCH}$$

where $\alpha_1^{(i)} = \alpha = 0.8$ under the null and $\alpha_1^{(i)} = 0.5 + 0.3 \cdot I_{\{i > \lfloor nt^* \rfloor\}}$ for $t^* \in \{0.25, 0.5, 0.75\}$ under the alternative. Note that, by (5.19), varying $\alpha_1^{(i)}$ changes the tail dependence coefficient. When fitting a GARCH(1,1)-model to empirical data one typically finds that $\alpha_1 + \beta$ is close to, but smaller than one, which motivated our choice for α_1 and β under the null. We try to detect a change in the probability of joint high-threshold exceedances of $\boldsymbol{V}_i = (|X_i|, |X_{i-1}|)$. Table 5.2 reports the results. The test is oversized for $n = 500$ and only marginally so for $n = 2000$. Remarkably, this holds irrespective of the choice of k, which we varied much more widely here than is common in extreme value theory to demonstrate the robustness of our results. Power seems to be a convex function in k with a peak between around $k/n = 0.3$. Quite expectedly, power increases in n. Late breaks ($t^* = 0.75$) are more frequently detected than early breaks ($t^* = 0.25$), with breaks in the middle being most frequently detected. Of course, by varying $\alpha_1^{(i)}$ we have also changed the marginal distribution so that, strictly speaking, the result of Section 5.2.3 does not apply. Nonetheless, the power of our test is very satisfactory.

Furthermore, we investigate the multivariate stochastic volatility (MSV) model from Example 5.2, where q, $h_{1,i}$ and $h_{2,i}$ are all generated according to zero-mean AR(1)-models with autoregressive parameters set to 0.3. For the innovations $\boldsymbol{Z}_i$ we chose the bivariate t-distribution with density

$$\boldsymbol{x} \mapsto c(1 + \boldsymbol{x}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x}/\nu)^{-(\nu+2)/2}.$$

This is a multivariate regularly varying distribution as required (cf., e.g., Resnick, 2007, Example 6.2). We choose a fairly heavy-tailed $\nu = 5$ and for the dispersion matrix

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} 1 & -0.3 - \delta \cdot I_{\{i > \lfloor nt^* \rfloor\}} \\ -0.3 - \delta \cdot I_{\{i > \lfloor nt^* \rfloor\}} & 1 \end{pmatrix},$$

where $\delta = 0$ under the null and $\delta = 0.6$ under the alternative. Note that by (5.23) this changes the tail dependence coefficient of the model, yet does not change the marginal distributions of $\boldsymbol{V}_i = (X_i, Y_i)$, because the diagonal elements of $\boldsymbol{\Sigma}_i$ are kept constant. The results (also displayed in Table 5.2) are largely similar. Here, size distortions are generally smaller.

DGP	Hyp.	t^*	$n = 500$					$n = 2000$				
			k/n									
			0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
(GARCH)	$\mathcal{H}_0$		7.7	7.4	7.2	7.8	5.9	6.0	6.6	5.9	5.8	6.2
	$\mathcal{H}_1$	0.25	11	14	15	16	13	30	41	45	44	44
		0.50	25	29	30	29	24	77	82	82	78	74
		0.75	26	25	24	23	17	66	66	63	56	52
(MSV)	$\mathcal{H}_0$		6.5	6.2	6.4	6.5	5.6	5.5	5.5	5.8	5.0	5.5
	$\mathcal{H}_1$	0.25	13	20	22	21	12	39	64	75	72	50
		0.50	18	32	40	38	22	59	88	95	94	74
		0.75	10	17	23	24	13	32	62	77	75	48

Table 5.2: Rejection frequencies at 5% of test based on $\mathcal{T}_n$ for n realizations of (GARCH) and (MSV)

Note that for detecting breaks in the serial dependence structure as in the the DGP in (GARCH) the $\mathcal{S}_n$-test is not suitable, because it is based on (roughly i.i.d.) estimated residuals. So there is no change in the dependence structure to be detected here and, additionally, the residuals are (roughly) asymptotically independent in the sense of (5.9), failing Assumption 3.1 in Bücher *et al.* (2015) (see also Remark 5.4 (a)). Simulating anyway, the $\mathcal{S}_n$-test is very oversized, invalidating test results.

As for the MSV-model it was not possible to fit a standard univariate model to the second component Y_i (e.g., GARCH(1,1) estimates did not converge properly). Hence, it is unclear how to proceed with the $\mathcal{S}_n$-test in this case, which could possibly be a frequent occurrence. Nonetheless, we would like to compare our test with the $\mathcal{S}_n$-test. The setup for the $\mathcal{S}_n$-test in the model (5.24) below is optimal - we suppose complete knowledge of the model and, what is even more, we suppose that the break in the tail dependence of the observed time series is only due to a break in the tail dependence of the innovations driving the time series. The latter is of course another nontrivial restriction. Indeed, the following example constructs a bivariate stochastic volatility model, whose tail dependence is determined by that of the volatility and the innovation process. Detecting breaks in such models will yet again be a daunting test for an innovation-based test.

Example 5.3. Consider

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,i} & 0 \\ 0 & \varepsilon_{2,i} \end{pmatrix} \begin{pmatrix} \sigma_{1,i} \\ \sigma_{2,i} \end{pmatrix} =: \boldsymbol{A}_i \boldsymbol{\sigma}_i,$$

where $\boldsymbol{\sigma}_i$ is strictly stationary and multivariate regular varying on $[-\infty,\infty]^2 \setminus \{\boldsymbol{0}\}$ with index $\alpha > 0$ and limit measure $\mu\{\cdot\}$. The $\varepsilon_{m,i}$ $(m = 1, 2)$ are i.i.d. independently of each other and of $\{(\sigma_{1,i}, \sigma_{2,i})'\}$ with $\mathbb{E}|\varepsilon_{m,i}|^{\alpha+\delta} < \infty$ for some $\delta > 0$. Then, by Basrak *et al.* (2002, Cor. A.2), $(X_i, Y_i)'$ is also multivariate regularly varying with index α and limit measure

$$\widetilde{\mu}\{\cdot\} = \mathbb{E}\mu\{\boldsymbol{A}_0^{-1}(\cdot)\},$$

i.e., the heavy tails are inherited from the volatility process and the tail dependence is determined by both the volatility (through $\mu\{\cdot\}$) and the innovation process (through the matrix $\boldsymbol{A}_i^{-1}$ in $\widetilde{\mu}$). Hence, a break in the extremal dependence of $(X_i, Y_i)'$ may not be picked up with a residual-based test, because it is due to a change in the volatility process $\boldsymbol{\sigma}_i$.

We use the model from Bücher *et al.* (2015, Eq. (16)):

$$\begin{cases} X_i = \sigma_{i,1} Z_{i,1}, & \sigma_{i,1}^2 = \alpha_{0,1} + \alpha_{1,1} X_{i-1}^2 + \beta_1 \sigma_{i-1,1}^2, \\ Y_i = \sigma_{i,2} Z_{i,2}, & \sigma_{i,2}^2 = \alpha_{0,2} + \alpha_{1,2} Y_{i-1}^2 + \beta_2 \sigma_{i-1,1}^2, \end{cases} \tag{5.24}$$

where we chose the same parameters as for model (GARCH) under the alternative, i.e.,

$$\alpha_{0,1} = \alpha_{0,2} = 0.01, \quad \alpha_{1,1} = 0.5, \quad \alpha_{1,2} = 0.8, \quad \beta_1 = \beta_2 = 0.15.$$

Let the bivariate noise $(Z_{i,1}, Z_{i,2})'$ have a Clayton copula with parameter $\theta = \theta_i$ and marginal distributions given by $Z_{i,1} \sim \mathcal{N}(0,1)$ and $\sqrt{3} Z_{i,2} \sim t_3$. See Bücher *et al.* (2015, Sec. 4.1) for details on the copula and a method for simulation. Note that the Clayton copula generates lower tail dependent data with upper tail *in*dependence. So before applying our test we premultiply the bivariate random vectors by -1 to obtain upper tail dependent data as required for our test. Of course a power comparison of both tests can only be meaningful when they have roughly the same size under the null. Hence we also run a simulation under the null with $\theta = 1$ and one under the alternative with $\theta = \theta_i = 1 + 2 \cdot I_{\{i > \lfloor nt^* \rfloor\}}$. The results are shown in Table 5.3.

For our test the results are qualitatively similar to those for (MSV). As already found in Bücher *et al.* (2015, Sec. 4.2), the $\mathcal{S}_n$-test is a bit undersized. Under the alternative it always achieves the highest power when $k/n = 0.3$ while ours does so for $k/n = 0.4$. Comparing the numbers for the rejection frequencies, we see that the $\mathcal{T}_n$-test almost always has higher power. In view of the lower size of the $\mathcal{S}_n$-test we conclude that even if all assumptions for this test to work 'properly' are satisfied (i.e., break only in the innovations and perfect knowledge of the model) the advantage over our more robust suggestion in this situation seems modest, if it exists at all. We offer the following explanation for this result, that may be surprising in view of the putative superiority of parametric test. Consider again the DGP in (5.24). Say that, as under our alternative, after the break the innovation vector $(Z_{i,1}, Z_{i,2})'$ became more tail dependent. But then not only through this will X_i and Y_i become more dependent, but also through the volatilities $\sigma_{i,1}$ and $\sigma_{i,2}$, which of course contain the now more dependent X_i and Y_i in their equations. So only focusing on the noise will neglect the additional effect of more dependent volatilities here.

Test	Hyp.	t^*	$n = 500$					$n = 2000$				
			k/n									
			0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
$\mathcal{T}_n$	$\mathcal{H}_0$		6.8	6.1	6.8	7.5	5.9	5.3	5.4	5.6	5.2	5.6
	$\mathcal{H}_1$	0.25	10	13	16	18	17	20	39	52	58	58
		0.50	15	23	29	33	29	37	66	80	85	84
		0.75	13	17	20	22	18	26	45	57	60	60
$\mathcal{S}_n$	$\mathcal{H}_0$		4.4	3.3	2.3	1.9	1.4	4.1	3.3	2.6	2.4	1.6
	$\mathcal{H}_1$	0.25	8.9	12	12	10	7.0	30	46	51	48	36
		0.50	17	23	25	21	16	55	76	80	80	68
		0.75	12	15	16	12	8.7	39	55	59	54	42

Table 5.3: Rejection frequencies at 5% of tests based on $\mathcal{T}_n$ and $\mathcal{S}_n$ developed in Bücher *et al.* (2015) for n realizations of (5.24)

Overall, our simulations demonstrate good size of our test even for $n = 500$ and very satisfactory power for $n = 2000$. Results under the null are very insensitive to the choice of k, while the highest power is achieved by choosing $k/n \approx 0.3$. Furthermore, our test can be applied in a range of settings, where it is difficult to apply the $\mathcal{S}_n$-test. When both tests are reasonable, the power advantage of the $\mathcal{S}_n$-test seems mild, if it exists at all.

5.4 Applications

In this section we apply our test in two settings. One where we test for breaks in the extremal dependence structure between two asset returns (Section 5.4.1) and another where we investigate the PA-extremogram, i.e., a serial extremal dependence measure (Section 5.4.2).

5.4.1 A financial crisis example

We examine whether the extremal dependence of log-returns between the indices S&P 500 and the DAX changed during the financial crisis of 2007-08. The question is of obvious relevance to investors - spreading assets across regions is beneficial to portfolios because of its diversifying effect and, as the saying has it, diversification is the only free lunch in finance. If however asset prices start to co-move (across asset classes or regions), then diversification will be less of a free lunch and a rebalancing of the portfolio may be required. As already mentioned, this phenomenon is called 'diversification meltdown' in the literature. The recent financial crisis was not a one-day event (like Black Monday, which was investigated in Bücher *et al.* (2015, Sec. 5.2)) but played out over several years, such that one has no strong prior beliefs on the location of a possible change point. Hence, testing for a change with an unknown location is required here. We analyze both upper and lower tail dependence separately and come to similar conclusions.

We use log-returns based on adjusted daily closing prices from 2004 to 2011 based on both indices and kept those where data for both are available, leaving us with a total of 1999 observations. Throughout we choose $r_n = \left\lfloor 15 \cdot k^{0.2} \right\rfloor$ as recommended in the previous section. The two time series are displayed in panels (a) and (b) of Figure 5.1. The scatter plots of the S&P 500 and DAX log-returns before and after investment bank Lehman Brothers filed for bankruptcy protection on September 15th, 2008 are shown in panels (c) and (d). The date is chosen somewhat arbitrarily, yet it was the largest bankruptcy (by assets) during the recent financial crisis and furthermore marked the beginning of the wildest swings in returns (panels (a) and (b)). Panel (c) clearly shows that there is some mild clustering of extremes in the upper-right and lower-left quadrant. This effect is much more pronounced in panel (d), giving a first indication of a break in the dependence structure.

More formally, we apply our test. Before doing so we check that our test may reasonably be applied. In view of the large sample size we may fit an ARMA(p_1, q_1)-GARCH(p_2, q_2)-model with $p_i, q_i \leq 2$ and select the one with the lowest AIC. This suggests an ARMA(1,2)-GARCH(2,2) for the S&P 500 and an ARMA(0,2)-GARCH(2,1) for the DAX data. Ljung-Box tests on the raw and squared standardized residuals indicate that either time series may be well-described by these stationary and heavy-tailed models.

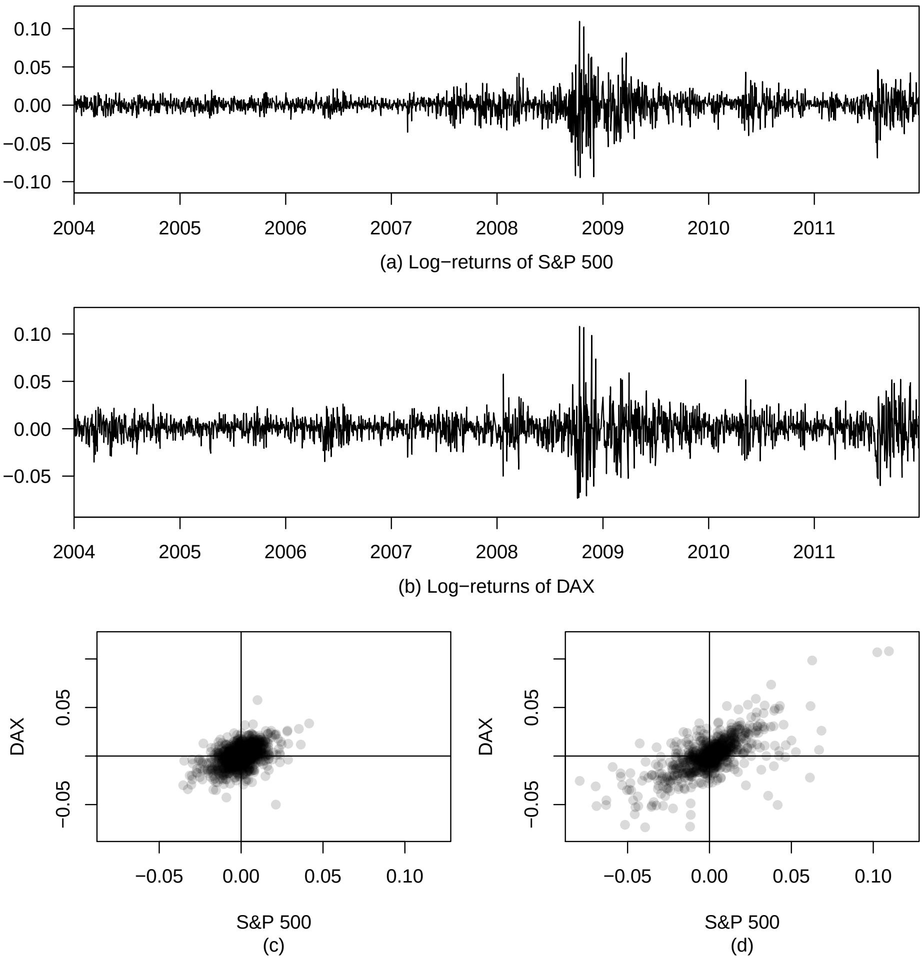

Figure 5.1: Plot of log-returns of S&P 500 in (a) and DAX in (b). Scatter plots of DAX and S&P 500 log-returns before Lehman bankruptcy on September 15th, 2008 in (c) and after in (d).

Figure 5.2 plots the values of the test statistic $\mathcal{T}_n$ for different values of k/n in for the log-returns (solid line) and for the (positive) log-losses (dashed line) of both series. The test statistic for the null of constant lower tail dependence is above the 99%-critical value for all reasonable values of k/n between 0.1 and 0.4. The test statistic for constant upper tail dependence yields even more significant results in that k/n-region. In the simulations power was seen to decline when k/n approached 0.5, which may also explain the downward slope in Figure 5.2. Overall, this is convincing evidence for a break in the lower and, even more, in the upper tail dependence of the bivariate time series.

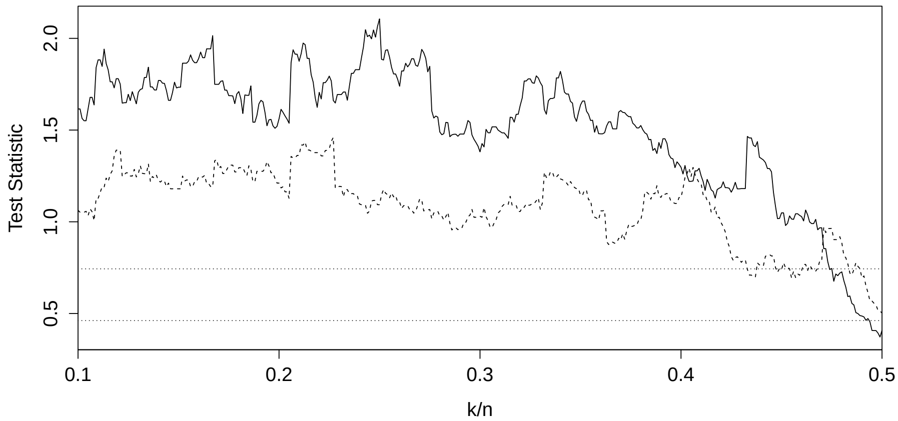

Figure 5.2: Plot of $k/n \mapsto \mathcal{T}_n$ for log-returns (solid) and log-losses (dashed). 95%- and 99%-critical values (dotted horizontal lines).

Two questions remain: first, in which direction was the break, i.e., are extremal co-movements more or less likely after the break? Second, when did the break(s) occur? Figure 5.3, which plots the normalized $G_n(t)$, provides some evidence. Recall that under the null the sample path of $G_n(\cdot)/\widehat{\sigma}$ should look like that of a Brownian bridge. The plot for both tests is almost exclusively negative, so by the definition of $G_n(t)$ in (5.12) extremal dependence in the upper and the lower tail between S&P 500 and DAX log-returns has likely intensified during the crisis, which is evidence for 'diversification meltdown'. Furthermore, the minimum of $t \mapsto G_n(t)/\widehat{\sigma}$ is attained on October 31, 2008 for the solid line and July 15, 2007 for the dashed line. Splitting the bivariate sample at both minima and testing for another break in the (upper and lower) extremal dependence in the respective subsamples, we found no evidence of further breaks. All in all, this suggests a break somewhere around the beginning or middle of the crisis after which extremal co-movements in both indices have become more frequent.

5.4.2 Confidence bounds for the PA-extremogram

We reconsider the (positive) FTSE log-losses $X_1, \ldots, X_{6653}$ from April 4, 1984 to October 2, 2009 (downloaded from *finance.yahoo.com*), already investigated in Davis *et al.* (2012, Sec. 3.1). We do so to compare the method of computing confidence bounds using Corollary 5.1 with the bootstrapped-based one of Davis *et al.* (2012).

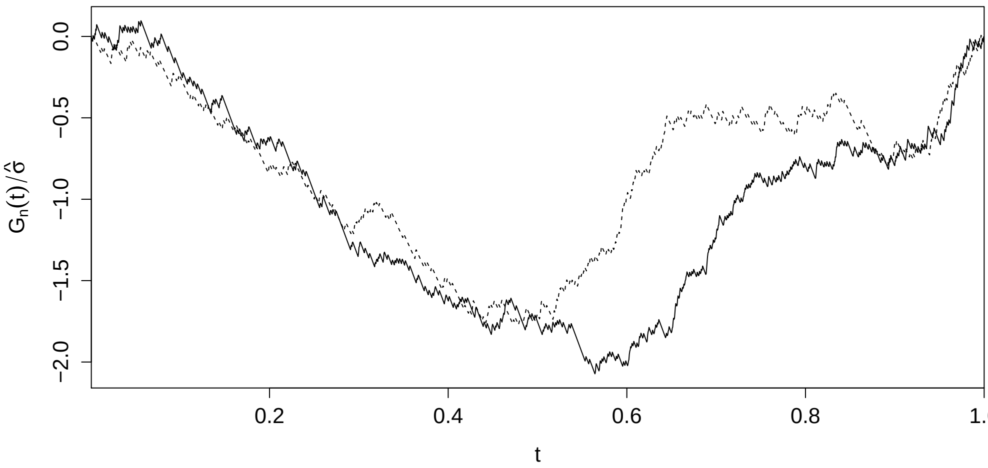

Figure 5.3: Plot of $t \mapsto G_n(t)/\widehat{\sigma}$ for $k/n = 0.3$ for log-returns (solid) and log-losses (dashed).

As in Davis *et al.* (2012, Sec. 3.1) we compute the PA-extremogram along with 95%-confidence bounds for $\boldsymbol{V}_i = (X_i, X_{i-h})'$, $h = 1, \ldots, 40$. The results are displayed in Figure 5.4. Comparing these with Davis *et al.* (2012, Fig. 3.1) we see that our method yields similar, yet slightly wider confidence bounds, for both the raw log-losses and the standardized GARCH(1,1)-residuals (using t-noise). The qualitative conclusions are the same: panel (a) suggests the presence of extremal serial dependence in the log-losses, while we can infer from panel (b) that this extremal serial dependence is well-captured by a GARCH(1,1)-specification, as the residuals exhibit roughly the behavior expected of i.i.d. data.

Since the time period spans over 25 years (and particularly the rise of computerized trading), it is certainly worthwhile to test for structural breaks in the extremal serial dependence. Only for 5 out of the 40 lags do we find insignificant results at the 1%-level, indicating that changes have indeed occurred at most lags. This of course sheds some doubt on the estimates and confidence bounds of the PA-extremogram and highlights the importance of testing for breaks as a pre-step.

So we suggest the following integrated method for estimation and inference for the PA-extremogram. First, apply the structural break test based on $\mathcal{T}_n$. Then, if no change is detected, Corollary 5.1 provides a consistent estimator along with central limit theory for inference.

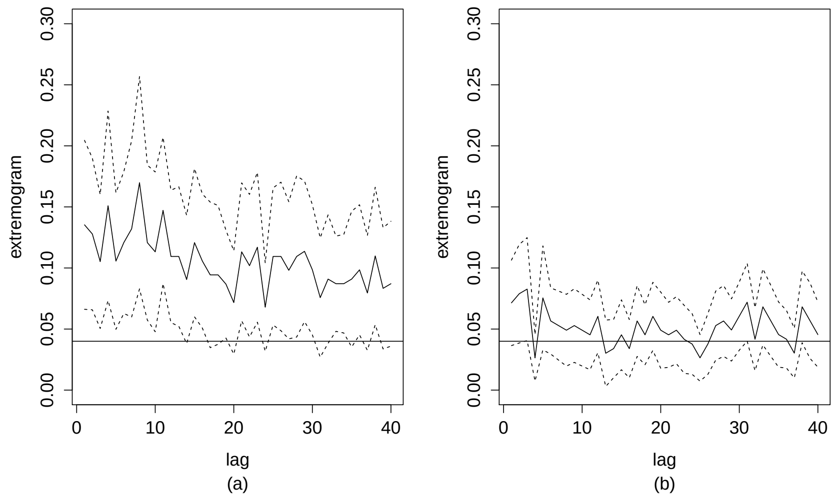

Figure 5.4: Plot of PA-extremogram (jagged solid line) with 95%-confidence interval (dashed lines) for FTSE log-losses in (a) and standardized GARCH(1,1)-residuals in (b). Straight solid line at 0.04 indicates value of PA-extremogram for i.i.d. data.

5.5 Proofs

In the following K and $\widetilde{\delta}$ denote large and small positive constants that may change from line to line. As usual, define $\|\cdot\|_p$ to be the L_p-norm, i.e., $\|X\|_p := \left[\mathbb{E}\,|X|^p\right]^{1/p}$, $|\cdot|$ applied to a set the cardinality, $\stackrel{\mathcal{D}}{=}$ equality in distribution and $a_n \sim b_n$ asymptotic equivalence, i.e., $\lim_{n\to\infty} a_n/b_n = 1$.

For $\boldsymbol{\xi} = (\xi_1, \xi_2)' \in \mathbb{R}^2$ put

$$A_i(\boldsymbol{\xi}) := I_{\left\{\log(X_i/b_x) > \xi_1/\sqrt{k}, \log(Y_i/b_y) > \xi_2/\sqrt{k}\right\}} \qquad \text{and} \qquad \widetilde{A}_i(\boldsymbol{\xi}) := A_i(\boldsymbol{\xi}) - \mathbb{E}[A_i(\boldsymbol{\xi})],$$

where here and in the following $b_z := b_z(n/k)$, $z \in \{x, y\}$, for short.

Lemma 5.1. *Suppose assumption* **(D3)** *is met. Then uniformly on every compact ξ-set*

$$\mathrm{P}\left\{\log(Z_i/b_z) > \xi/\sqrt{k}\right\} = \frac{k}{n}\left(1 - \alpha_z \frac{\xi}{\sqrt{k}} + o\left(\frac{1}{\sqrt{k}}\right)\right).$$

Proof. Follow the steps in the proof of Theorem 2.4 in Hsing (1991) to the last display on p. 1553 and combine with the fact that (5.5) and (5.6) hold uniformly on compact λ-sets in $(0, \infty)$ (cf. Hsing, 1991, Theorem 2.3). $\square$

Proof of Proposition 5.1: Using stationarity we get

$$
\begin{aligned}
& \frac{n}{r_n k} \operatorname{Var}\left(\sum_{i=1}^{r_n} I_{\left\{X_i>b_x, Y_i>b_y\right\}}\right) \\
= \; & \frac{n}{r_n k}\left[\sum_{i=1}^{r_n} \operatorname{Var}\left(I_{\left\{X_i>b_x, Y_i>b_y\right\}}\right)\right. \\
& \left. \quad +2 \sum_{m=1}^{r_n}\left(r_n-m\right) \operatorname{Cov}\left(I_{\left\{X_1>b_x, Y_1>b_y\right\}}, I_{\left\{X_{1+m}>b_x, Y_{1+m}>b_y\right\}}\right)\right] \\
=: \; & A_n+B_n .
\end{aligned}
$$

By (5.7) and Lemma 5.1 we have

$$
A_n=\frac{n}{k}\left[\mathrm{P}\left\{X_i>b_x, Y_i>b_y\right\}-\mathrm{P}\left\{X_i>b_x\right\} \mathrm{P}\left\{Y_i>b_y\right\}\right] \underset{(n \rightarrow \infty)}{\longrightarrow} c_0 .
$$

Furthermore, using Lemma 5.1 again

$$
\begin{aligned}
& \operatorname{Cov}\left(I_{\left\{X_1>b_x, Y_1>b_y\right\}}, I_{\left\{X_{1+m}>b_x, Y_{1+m}>b_y\right\}}\right) \\
= & \mathrm{P}\left\{X_1>b_x, Y_1>b_y, X_{1+m}>b_x, Y_{1+m}>b_y\right\}-\mathrm{P}\left\{X_1>b_x, Y_1>b_y\right\}^2 \\
= & \mathrm{P}\left\{X_1>b_x, Y_1>b_y, X_{1+m}>b_x, Y_{1+m}>b_y\right\}-\mathcal{O}\left((k / n)^2\right),
\end{aligned}
$$

so that with (5.7), (5.8) and $\lim_{n \rightarrow \infty} r_n k / n=0$

$$
\begin{aligned}
B_n= & 2 \frac{n}{r_n k} \sum_{m=1}^h\left(r_n-m\right) \mathrm{P}\left\{X_1>b_x, Y_1>b_y, X_{1+m}>b_x, Y_{1+m}>b_y\right\} \\
& +\sum_{m=h+1}^{r_n}\left(r_n-m\right) \mathrm{P}\left\{X_1>b_x, Y_1>b_y, X_{1+m}>b_x, Y_{1+m}>b_y\right\}+o(1) \\
\underset{(n \rightarrow \infty)}{\longrightarrow} & 2 \sum_{m=1}^{\infty} c_m .
\end{aligned}
$$

as $n \rightarrow \infty$ followed by $h \rightarrow \infty$. The result follows. □

The road map for the proof of Theorem 5.1 is as follows. In a first step we consider convergence of

$$\sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[A_i(\mathbf{0})-\left(\frac{k}{n}\right)^2-\left[\mathrm{P}\left\{X_i>b_x,Y_i>b_y\right\}-\mathrm{P}\left\{X_i>b_x\right\}\mathrm{P}\left\{Y_i>b_y\right\}\right]\right]\right\}. \tag{5.25}$$

Since, by Lemma 5.1 and the assumed stationarity of the marginals even under the alternative,

$$\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[\left(\frac{k}{n}\right)^2-\mathrm{P}\left\{X_i>b_x\right\}\mathrm{P}\left\{Y_i>b_y\right\}\right]=o\left(\frac{1}{\sqrt{k}}\right)\quad \text{uniformly in } t\in[0,1],$$

it suffices to consider the non-negligible parts in (5.25). In a second step, we show that b_x, b_y in $A_i(\mathbf{0})$ may be replaced by suitable empirical counterparts.

The first step is taken care of by

Lemma 5.2. *Suppose assumptions* **(D1)**-**(D3)** *are met. Then under $\mathcal{H}_0$ and $\mathcal{H}_1$*

$$\sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\widetilde{A}_i(\mathbf{0})\right\}\xrightarrow[(n\to\infty)]{\mathcal{D}}\sigma W(t)\qquad \text{in } D[0,1],$$

where $\{W(t)\}_{t\in[0,1]}$ denotes a standard Brownian motion.

Proof. The rough outline of the proof is as the one of Theorem 2.4. For $t\in[0,1]$ define

$$m_n(t):=\left\lfloor\frac{\lfloor nt\rfloor}{r_n+l_n}\right\rfloor$$

and for $j=1,\dots,m_n(1)$ define B_j (the big blocks) and S_j (the small blocks) to be consecutive blocks of integers of length $|B_j|=r_n$ and $|S_j|=l_n$, i.e.,

$$\begin{aligned}&B_1=\{1,\dots,r_n\},\ S_1=\{r_n+1,\dots,r_n+l_n\},\\ &\qquad B_2=\{r_n+l_n+1,\dots,2r_n+l_n\},\ S_2=\{2r_n+l_n+1,\dots,2r_n+2l_n\},\ \text{etc.}\end{aligned}$$

Choose the length of $B_{m_n(t)+1}$ such that the integers $\{1,\dots,\lfloor nt\rfloor\}$ are covered. Now decompose

$$\sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\widetilde{A}_i(\mathbf{0})\right\}=\sum_{j=1}^{m_n(t)}Y_j^B+\sum_{j=1}^{m_n(t)}Y_j^S+R_n(t)$$

where

$$
\begin{aligned}
Y_j^B &:= Y_{j,n}^B := \frac{1}{\sqrt{k}} \sum_{i \in B_j} \widetilde{A}_i(\mathbf{0}), \\
Y_j^S &:= Y_{j,n}^S := \frac{1}{\sqrt{k}} \sum_{i \in S_j} \widetilde{A}_i(\mathbf{0}), \\
R_n(t) &:= \frac{1}{\sqrt{k}} \sum_{i \in B_{m_n(t)+1}} \widetilde{A}_i(\mathbf{0}).
\end{aligned}
\tag{5.26}
$$

We will consider these terms separately. First, noting that $|B_{m_n(t)+1}| \leq r_n + l_n - 1$,

$$
0 \leq \sup_{t \in [0,1]} \left| R_n(t) \right| \leq 2 \frac{r_n + l_n - 1}{\sqrt{k}} \underset{(n \to \infty)}{\overset{\textbf{(D1)}}{\longrightarrow}} 0. \tag{5.27}
$$

Let $\widetilde{Y}_j^B \overset{\mathcal{D}}{=} Y_j^B$ be independent copies of the Y_j^B's and likewise for the Y_j^S's. By generalizations of the arguments in the proof of Theorem 2.4 to the array case (see, e.g., Drees and Rootzén, 2010, Proof of Lemma 5.1 for the non-functional case) $\sum_{j=1}^{m_n(t)} Y_j^B$ and $\sum_{j=1}^{m_n(t)} Y_j^S$ have the same weak limit in $D[0,1]$ as

$$
\sum_{j=1}^{m_n(t)} \widetilde{Y}_j^B \qquad \text{and} \qquad \sum_{j=1}^{m_n(t)} \widetilde{Y}_j^S \tag{5.28}
$$

if the respective limits exist. To derive the weak limit of $\sum_{j=1}^{m_n(t)} \widetilde{Y}_j^B$ we verify that the conditions of the martingale difference array functional central limit theorem given in Gaenssler and Haeusler (1986, Thm. 2.2) are satisfied. Using **(D2)** we get for all $t \in [0,1]$

$$
\sum_{j=1}^{m_n(t)} \operatorname{Var}(\widetilde{Y}_j^B) = \left\lfloor \frac{\lfloor nt \rfloor}{r_n + l_n} \right\rfloor \operatorname{Var}(Y_1^B) + o(1) \underset{(n \to \infty)}{\longrightarrow} \sigma^2 t
$$

and, because $\left|\widetilde{Y}_j^B\right| \leq 2 r_n / \sqrt{k} \to 0$ for all $j \geq 1$,

$$
\sum_{j=1}^{m_n(t)} \mathbb{E}\left[(\widetilde{Y}_j^B)^2 I_{\left\{ \left| \widetilde{Y}_j^B \right| \geq \eta \right\}} \right] \underset{(n \to \infty)}{\longrightarrow} 0 \qquad \forall\, \eta > 0
$$

is trivial. Hence, it follows that

$$\sum_{j=1}^{m_n(t)} Y_j^B \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \sigma W(t) \qquad \text{in } D[0,1]. \tag{5.29}$$

It remains to show negligibility of $\sum_{j=1}^{m_n(t)} \widetilde{Y}_j^S$. For this observe that by the c_r-inequality

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{l_n} I_{\{X_i>b_x,Y_i>b_y\}}\right) &\leq \mathbb{E}\left[\sum_{i=1}^{l_n} I_{\{X_i>b_x,Y_i>b_y\}}\right]^2 \\ &\leq l_n \sum_{i=1}^{l_n} \mathbb{E}\left[I_{\{X_i>b_x,Y_i>b_y\}}\right]^2 = \mathcal{O}\left(l_n^2 \frac{k}{n}\right)\end{aligned}$$

where the last equality follows from **(D2)** and Lemma 5.1. Applying Kolmogorov's inequality we thus get

$$\begin{aligned}\mathrm{P}\left\{\sup_{t\in[0,1]}\left|\sum_{j=1}^{m_n(t)} \widetilde{Y}_j^S\right| \geq \varepsilon\right\} &= \mathrm{P}\left\{\max_{1\leq l\leq m_n(1)}\left|\sum_{j=1}^{l} \widetilde{Y}_j^S\right| \geq \varepsilon\right\} \\ &\leq \varepsilon^{-2}\sum_{j=1}^{m_n(1)} \operatorname{Var}(\widetilde{Y}_j^S) \\ &\leq K\frac{n}{r_n k}\operatorname{Var}\left(\sum_{i=1}^{l_n} I_{\{X_i>b_x,Y_i>b_y\}}\right) = \mathcal{O}(l_n^2/r_n),\end{aligned}$$

which tends to zero as $n\to\infty$ by **(D1)**. □

In the second step we need to justify the replacement of b_x and b_y with $X_{(k+1)}$ and $Y_{(k+1)}$ in $I_{\{X_i>b_x,Y_i>b_y\}}$.

Lemma 5.3. *Suppose assumption* **(D3)** *is met. Then uniformly on every compact* $\boldsymbol{\xi}$*-set and uniformly in* $i\in\mathbb{N}$,

$$\mathbb{E}|A_i(\mathbf{0}) - A_i(\boldsymbol{\xi})|^r \leq c_r \max(\alpha_X,\alpha_Y)\frac{\sqrt{k}}{n}\left(|\xi_1|+|\xi_2|+o(1)\right),$$

where $r\geq 1$ *and* $c_r = 2^{r-1}$.

Proof. Lemma 5.1 implies uniformly in $i \in \mathbb{N}$ (recall the stationarity of the marginals)

$$\mathrm{P}\left\{0 < \log(X_i/b_x) \leq \xi_1/\sqrt{k}\right\} = \frac{\sqrt{k}}{n}\left(\alpha_x \xi_1 + o(1)\right), \quad \xi > 0,$$
$$\mathrm{P}\left\{\xi_1/\sqrt{k} < \log(X_i/b_x) \leq 0\right\} = \frac{\sqrt{k}}{n}\left(-\alpha_x \xi_1 + o(1)\right), \quad \xi < 0,$$

such that for any $\xi_1 \in \mathbb{R}$

$$\mathrm{P}\left\{\min(0,\xi_1/\sqrt{k}) < \log(X_i/b_x) \leq \max(0,\xi_1/\sqrt{k})\right\} = \frac{\sqrt{k}}{n}\left(\alpha_x |\xi_1| + o(1)\right), \tag{5.30}$$

where all $o(1)$-terms are uniform on compact ξ-sets. This relation also holds mutatis mutandis for Y_i. Now by the c_r-inequality

$$\begin{aligned}
&\mathbb{E}[A_i(\mathbf{0}) - A_i(\boldsymbol{\xi})]^r \\
&\leq \mathbb{E}\left[I_{\left\{\min(0,\xi_1/\sqrt{k})<\log(X_i/b_x)\leq\max(0,\xi_1/\sqrt{k})\right\}} + I_{\left\{\min(0,\xi_2/\sqrt{k})<\log(Y_i/b_y)\leq\max(0,\xi_2/\sqrt{k})\right\}}\right]^r \\
&\leq c_r \left[\mathrm{P}\left\{\min(0,\xi_1/\sqrt{k}) < \log(X_i/b_x) \leq \max(0,\xi_1/\sqrt{k})\right\}\right. \\
&\qquad\qquad \left. + \mathrm{P}\left\{\min(0,\xi_2/\sqrt{k}) < \log(Y_i/b_y) \leq \max(0,\xi_2/\sqrt{k})\right\}\right].
\end{aligned}$$

The results follows with (5.30) and its analogue for Y_i. □

Lemma 5.4. *Suppose assumptions* **(D1)** *and* **(D3)** *are met. Then for every* $\boldsymbol{\xi} = (\xi_1, \xi_2)' \in \mathbb{R}^2$

$$\sup_{t\in[0,1]} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \widetilde{A}_i(\mathbf{0}) - \widetilde{A}_i(\boldsymbol{\xi}) \right| = o_{\mathrm{P}}(1).$$

Proof. The arguments resemble those in the proof of Lemma 5.2. Write

$$\frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \widetilde{A}_i(\mathbf{0}) - \widetilde{A}_i(\boldsymbol{\xi}) = \sum_{j=1}^{m_n(t)} Y_j^B + \sum_{j=1}^{m_n(t)} Y_j^S + R_n(t),$$

where $Y_j^B, Y_j^S, R_n(t)$ are defined as in (5.26) with $\widetilde{A}_i(\mathbf{0})$ replaced by $\widetilde{A}_i(\mathbf{0}) - \widetilde{A}_i(\boldsymbol{\xi})$.

Considering the independent versions as in (5.28) again we get uniformly in $j \in \mathbb{N}$

$$\begin{aligned}
\operatorname{Var}(\widetilde{Y}_j^B) &\leq \frac{1}{k}\mathbb{E}\left[\sum_{i\in B_j} A_i(\mathbf{0}) - A_i(\boldsymbol{\xi})\right]^2 = \frac{1}{k}\left\{\left\|\sum_{i\in B_j} A_i(\mathbf{0}) - A_i(\boldsymbol{\xi})\right\|_2\right\}^2 \\
&\leq \frac{1}{k}\left\{r_n \max_{i\in\{1,\dots,n\}} \|A_i(\mathbf{0}) - A_i(\boldsymbol{\xi})\|_2\right\}^2 = \frac{r_n^2}{k}\max_{i\in\{1,\dots,n\}} \mathbb{E}[A_i(\mathbf{0}) - A_i(\boldsymbol{\xi})]^2 \\
&\leq 2\max(\alpha_x,\alpha_y)\frac{r_n^2}{n\sqrt{k}}(|\xi_1| + |\xi_2| + o(1)),
\end{aligned}$$

where we used Minkowski's inequality for the second inequality and Lemma 5.3 for the third. With this

$$\begin{aligned}
\mathrm{P}\left\{\sup_{t\in[0,1]}\left|\sum_{j=1}^{m_n(t)} \widetilde{Y}_j^B\right| \geq \varepsilon\right\} &= \mathrm{P}\left\{\max_{1\leq k\leq m_n(1)}\left|\sum_{j=1}^{k} \widetilde{Y}_j^B\right| \geq \varepsilon\right\} \\
&\leq \varepsilon^{-2}\sum_{j=1}^{m_n(1)} \operatorname{Var}(\widetilde{Y}_j^B) \\
&\leq K\varepsilon^{-2}\frac{n}{r_n + l_n}\frac{r_n^2}{n\sqrt{k}} \\
&= \mathcal{O}(r_n/\sqrt{k}) = o(1),
\end{aligned}$$

where we used Kolmogorov's inequality in the second step and Minkowski's inequality in the third. So in particular going back to the original probability space

$$\sum_{j=1}^{m_n(t)} Y_j^B \xrightarrow[(n\to\infty)]{\mathcal{D}} 0 \qquad \text{in } D[0,1],$$

whence with the CMT

$$\sup_{t\in[0,1]}\left|\sum_{j=1}^{m_n(t)} Y_j^B\right| \xrightarrow[(n\to\infty)]{\mathcal{D}} 0.$$

A similar argument proves that $\sup_{t\in[0,1]}\left|\sum_{j=1}^{m_n(t)} Y_j^S\right| \xrightarrow[(n\to\infty)]{\mathcal{D}} 0$. Uniform negligibility of $R_n(t)$ follows as in (5.27). $\square$

Lemma 5.5. *Suppose assumptions* **(D1)** *and* **(D3)** *are met. Then for every* $\boldsymbol{\xi} =$

$(\xi_1, \xi_2)', \boldsymbol{\nu} = (\nu_1, \nu_2)' \in [-K, K]^2$

$$\lim_{\rho \to 0} \limsup_{n \to \infty} \sup_{|\boldsymbol{\xi} - \boldsymbol{\nu}| \le \rho} \frac{n}{\sqrt{k}} \max_{i \in \{1, \dots, n\}} \mathbb{E} \left| A_i(\boldsymbol{\xi}) - A_i(\boldsymbol{\nu}) \right| = 0, \tag{5.31}$$

$$\lim_{\rho \to 0} \limsup_{n \to \infty} \mathrm{P} \left\{ \sup_{|\boldsymbol{\xi} - \boldsymbol{\nu}| \le \rho} \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \left| A_i(\boldsymbol{\xi}) - A_i(\boldsymbol{\nu}) \right| \ge \varepsilon \right\} = 0, \tag{5.32}$$

where $|\boldsymbol{\xi} - \boldsymbol{\nu}| := \max(|\xi_1 - \nu_1|, |\xi_2 - \nu_2|)$ *denotes the maximum norm in* $\mathbb{R}^2$*.*

Proof. For (5.31) we have uniformly in $\boldsymbol{\xi}, \boldsymbol{\nu}$ and i

$$\begin{aligned} \mathbb{E} \left| A_i(\boldsymbol{\xi}) - A_i(\boldsymbol{\nu}) \right| &\le \mathrm{P} \left\{ \min(\xi_1/\sqrt{k}, \nu_1/\sqrt{k}) < \log(X_i/b_x) \le \max(\xi_1/\sqrt{k}, \nu_1/\sqrt{k}) \right\} \\ &\quad + \mathrm{P} \left\{ \min(\xi_2/\sqrt{k}, \nu_2/\sqrt{k}) < \log(Y_i/b_y) \le \max(\xi_2/\sqrt{k}, \nu_2/\sqrt{k}) \right\} \\ &\le \max(\alpha_x, \alpha_y) \frac{\sqrt{k}}{n} \left(|\xi_1 - \nu_1| + |\xi_2 - \nu_2| + o(1) \right), \end{aligned}$$

where the last line follows similarly as (5.30).

To prove (5.32) assume without loss of generality (w.l.o.g.) that $K/\rho \in \mathbb{N}$. Note that

$$\sup_{|\boldsymbol{\xi} - \boldsymbol{\nu}| \le \rho} \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \left| A_i(\boldsymbol{\xi}) - A_i(\boldsymbol{\nu}) \right| \le \max_{(l_1, l_2)'} \frac{1}{\sqrt{k}} \sum_{i=1}^{n} A_i((l_1 + 2, l_2 + 2)'\rho) - A_i((l_1, l_2)'\rho),$$

where the maximum is taken over

$$\mathcal{L} := \left\{ (l_1, l_2)' \in \mathbb{Z}^2 \;\middle|\; [l_1, l_1 + 2] \times [l_2, l_2 + 2] \subset [-K/\rho, K/\rho]^2 \right\}.$$

Let $\varepsilon > 0$. Because of (5.31) there exist $\rho_0 = \rho_0(\varepsilon) > 0$ and $n_0 = n_0(\varepsilon)$ such that

$$\frac{n}{\sqrt{k}} \max_{i \in \{1, \dots, n\}} \mathbb{E} \left| A_i((l_1 + 2, l_2 + 2)'\rho) - A_i((l_1, l_2)'\rho) \right| < \frac{\varepsilon}{2} \qquad \forall\, \rho \in (0, \rho_0],\ n \ge n_0, \tag{5.33}$$

whence for $\rho \in (0, \rho_0]$

$$\begin{aligned} &\limsup_{n \to \infty} P \left\{ \max_{(l_1, l_2)'} \frac{1}{\sqrt{k}} \sum_{i=1}^{n} A_i((l_1 + 2, l_2 + 2)'\rho) - A_i((l_1, l_2)'\rho) > \varepsilon \right\} \\ &\quad \le \limsup_{n \to \infty} \sum_{(l_1, l_2)' \in \mathcal{L}} \mathrm{P} \left\{ \frac{1}{\sqrt{k}} \sum_{i=1}^{n} A_i((l_1 + 2, l_2 + 2)'\rho) - A_i((l_1, l_2)'\rho) > \varepsilon \right\} \end{aligned}$$

$$\overset{(5.33)}{\leq} \limsup_{n\to\infty} \sum_{(l_1,l_2)'\in\mathcal{L}} \mathrm{P}\left\{ \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \widetilde{A}_i((l_1+2,l_2+2)'\rho) - \widetilde{A}_i((l_1,l_2)'\rho) > \varepsilon/2 \right\} = 0$$

by Lemma 5.4. □

Lemma 5.6. *Suppose assumptions* **(D1)** *and* **(D3)** *are met. Then for every* $K > 0$

$$\sup_{\substack{t\in[0,1]\\ \boldsymbol{\xi}\in[-K,K]^2}} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt\rfloor} \widetilde{A}_i(\mathbf{0}) - \widetilde{A}_i(\boldsymbol{\xi}) \right| = o_{\mathrm{P}}(1). \tag{5.34}$$

Proof. Assume again w.l.o.g. that $K/\rho \in \mathbb{N}$. Put

$$\widetilde{\mathcal{L}} := \left\{ (l_1,l_2)' \in \mathbb{Z}^2 \;\middle|\; (l_1,l_2)' \in [-K/\rho, K/\rho]^2 \right\}.$$

Bound the left-hand side of (5.34) by

$$\max_{(l_1,l_2)'\in\widetilde{\mathcal{L}}} \sup_{t\in[0,1]} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt\rfloor} \widetilde{A}_i(\mathbf{0}) - \widetilde{A}_i((l_1,l_2)'\rho) \right| + \sup_{|\boldsymbol{\xi}-\boldsymbol{\nu}|\leq\rho} \sup_{t\in[0,1]} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt\rfloor} \widetilde{A}_i(\boldsymbol{\xi}) - \widetilde{A}_i(\boldsymbol{\nu}) \right|$$
$$=: A_n + B_n.$$

By Lemma 5.4 the term A_n tends to zero in probability. As for B_n we have

$$B_n \leq \sup_{|\boldsymbol{\xi}-\boldsymbol{\nu}|\leq\rho} \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \left| A_i(\boldsymbol{\xi}) - A_i(\boldsymbol{\nu}) \right| + \sup_{|\boldsymbol{\xi}-\boldsymbol{\nu}|\leq\rho} \frac{n}{\sqrt{k}} \max_{i\in\{1,\ldots,n\}} \mathbb{E}\left| A_i(\boldsymbol{\xi}) - A_i(\boldsymbol{\nu}) \right|.$$

By Lemma 5.5 the terms on the right-hand side tend to zero in probability as $n \to \infty$ and $\rho \to 0$. □

We state the next lemma in more generality than is strictly needed, yet the proof is more transparent this way. For $Z \in \{X, Y\}$ set

$$Z_{(k,t)} := (\lfloor kt\rfloor + 1)\text{-largest value of } Z_1, \ldots, Z_{\lfloor nt\rfloor}.$$

Lemma 5.7. *Suppose assumptions* **(D1)** *and* **(D3)** *are met. Then for* $Z \in \{X, Y\}$ *for any* $t_0 \in (0,1)$

$$\sup\left\{ \sqrt{k} \log(Z_{(k,t)}/b_z) \;\middle|\; t \in [t_0, 1] \right\} \underset{(n\to\infty)}{=} \mathcal{O}_{\mathrm{P}}(1).$$

Proof. W.l.o.g. we prove the result for $Z = X$ and define

$$B_i(\xi) := I_{\left\{\log(X_i/b_x) > \xi/\sqrt{k}\right\}} \qquad \text{and} \qquad \widetilde{B}_i(\xi) := B_i(\xi) - \mathbb{E}[B_i(\xi)].$$

For $\xi \in \mathbb{R}$

$$\begin{aligned}
\sqrt{k}\log(X_{(k,t)}/b_x) \le \xi \iff & \sum_{i=1}^{\lfloor nt \rfloor} I_{\left\{\log(X_i/b_x) > \xi/\sqrt{k}\right\}} \le \lfloor kt \rfloor \\
\iff & \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \widetilde{B}_i(\xi) \le \frac{\lfloor kt \rfloor}{\sqrt{k}} - \frac{\lfloor nt \rfloor}{\sqrt{k}} \mathbb{E}[A_i(\boldsymbol{\xi})] \\
\iff & \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \left(\widetilde{B}_i(\xi) - \widetilde{B}_i(0)\right) + \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \widetilde{B}_i(0) \le \alpha_x \xi t + o(1),
\end{aligned} \tag{5.35}$$

where we used Lemma 5.1 for the last equivalence. Since the first two sums on the left-hand side of (5.35) are $o_{\mathrm{P}}(1)$ uniformly in t by the same arguments used in the proof of Lemma 5.4 and the remaining term is $\mathcal{O}_{\mathrm{P}}(1)$ by the following argument, the result follows.

We need to prove that for all $\varepsilon > 0$ there exists a $K > 0$, such that

$$\mathrm{P}\left\{ \sup_{t \in [0,1]} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \widetilde{B}_i(0) \right| > K \right\} \le \varepsilon.$$

Use Markov's inequality to obtain

$$\begin{aligned}
\mathrm{P}\left\{ \sup_{t \in [0,1]} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \widetilde{B}_i(0) \right| > K \right\} &\le K^{-2} \mathbb{E}\left[\sup_{t \in [0,1]} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \widetilde{B}_i(0) \right| \right]^2 \\
&= K^{-2} \mathbb{E}\left[\max_{1 \le l \le n} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{l} \widetilde{B}_i(0) \right|^2 \right]
\end{aligned} \tag{5.36}$$

Now apply the mixingale inequality in Davidson (1994, Cor. 16.10). For this check that $\left\{\frac{1}{\sqrt{k}}\widetilde{B}_i(0)\right\}$ is indeed an $\mathcal{L}_2$-mixingale array of size $-1/2$ on the canonical filtration

$$\mathcal{F}_{n,i} := \sigma\left(\frac{1}{\sqrt{k}}\widetilde{B}_i(0), \frac{1}{\sqrt{k}}\widetilde{B}_{i-1}(0), \dots \right).$$

The second part of the definition in Davidson (1994, Def. 16.5) is trivially satisfied, while for the first we observe that by Davidson (1994, Thm. 14.2) for any $r \geq 2$

$$\begin{aligned}\left\|\mathbb{E}\left[\frac{1}{\sqrt{k}}\widetilde{B}_i(0)|\mathcal{F}_{n,i-l_n}\right]\right\|_2 &\leq 2(\sqrt{2}+1)[\beta_n(l_n)]^{1/2-1/r}\left\|\frac{1}{\sqrt{k}}\widetilde{B}_i(0)\right\|_r \\ &\leq K[\beta_n(l_n)]^{1/2-1/r}\frac{1}{\sqrt{k}}\left(P\left\{X_i > b_x\right\}\right)^{1/r} \\ &= \frac{K}{\sqrt{n}}\cdot\left[\frac{n}{k}\beta_n(l_n)\right]^{1/2-1/r} =: c_{n,i}\cdot\xi_{l_n}.\end{aligned}$$

By **(D1)** we have $\xi_{l_n} = \mathcal{O}(l_n^{-1/2+\varepsilon})$ for some $\varepsilon > 0$, such that the mixingale size is indeed $-1/2$ as required. Application of Davidson (1994, Cor. 16.10) hence yields

$$\mathbb{E}\left[\max_{1\leq l\leq n}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{l}\widetilde{B}_i(0)\right|^2\right] = \mathcal{O}\left(\sum_{i=1}^{n}c_{n,i}^2\right) = \mathcal{O}(1),$$

whence the result follows from (5.36). □

Proof of Theorem 5.1: Apply the CMT with $f(x(t)) = x(t) - tx(1)$ to Lemma 5.2 to obtain

$$\sqrt{k}t(1-t)\left(\frac{1}{kt}\sum_{i=1}^{\lfloor nt\rfloor}A_i(\mathbf{0}) - \frac{1}{k(1-t)}\sum_{i=\lfloor nt\rfloor+1}^{n}A_i(\mathbf{0})\right)\xrightarrow[(n\to\infty)]{\mathcal{D}}\sigma B(t) \qquad \text{in } D[0,1] \tag{5.37}$$

with $\{B(t)\}$ a Brownian bridge. Note that we have also used the null hypothesis and $\mathbb{E}[A_i(\mathbf{0})] = \mathcal{O}(k/n)$ by Lemma 5.1. Define

$$\boldsymbol{\xi}_n = \begin{pmatrix}\sqrt{k}\log(X_{(k+1)}/b_x)\\ \sqrt{k}\log(Y_{(k+1)}/b_y)\end{pmatrix}.$$

With this notation it suffices to show that

$$\sup_{t\in[0,1]}\left|\sqrt{k}\left(\frac{1-t}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[A_i(\boldsymbol{\xi}_n) - A_i(\mathbf{0})\right] - \frac{t}{k}\sum_{i=\lfloor nt\rfloor+1}^{n}\left[A_i(\boldsymbol{\xi}_n) - A_i(\mathbf{0})\right]\right)\right| \underset{(n\to\infty)}{=} o_{\mathrm{P}}(1). \tag{5.38}$$

We consider the left-hand side of (5.38) separately on the sets $\{|\boldsymbol{\xi}_n| \leq K\}$ and $\{|\boldsymbol{\xi}_n| > K\}$, where $|\cdot|$ again denotes the maximum norm in $\mathbb{R}^2$. On $\{|\boldsymbol{\xi}_n| \leq K\}$

we may bound the left-hand side of (5.38) by

$$2 \sup_{\substack{t\in[0,1] \\ \boldsymbol{\xi}\in[-K,K]^2}} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor nt \rfloor} \left[\widetilde{A}_i(\boldsymbol{\xi}) - \widetilde{A}_i(\mathbf{0}) \right] \right| = o_{\mathrm{P}}(1) \quad \text{by Lemma 5.6,}$$

where we also used Lemma 5.3 for the bound. For $\{|\boldsymbol{\xi}_n| > K\}$ we have, due to Lemma 5.7 for $t = 1$,

$$\mathrm{P}\left\{|\boldsymbol{\xi}_n| > K\right\} \leq \mathrm{P}\left\{|\sqrt{k}\log(X_{(k+1)}/b_x)| > K\right\} + \mathrm{P}\left\{|\sqrt{k}\log(Y_{(k+1)}/b_y)| > K\right\} \to 0,$$

as $n, K \to \infty$. Now (5.38) follows. □

Now we set out to prove Theorem 5.2. In a first step we show that $\widetilde{\sigma}^2 \xrightarrow[(n\to\infty)]{\mathrm{P}} \sigma^2$, where $\widetilde{\sigma}^2$ is defined in (5.14). Then, in a second step, we have to justify the replacement of b_x and b_y in $\widetilde{\sigma}^2$ by their empirical counterparts $X_{(k+1)}$ and $Y_{(k+1)}$. For all these steps it will be convenient to define

$$Y_{j,n}(\boldsymbol{\xi}) := \sum_{i=(j-1)r_n+1}^{jr_n} \frac{1}{\sqrt{k}} \widetilde{A}_i(\boldsymbol{\xi}), \qquad j = 1, \ldots, \lfloor n/r_n \rfloor .$$

The first step is completed by proving the following

Lemma 5.8. *Suppose assumptions* **(D1)-(D3)** *are met. Then under* $\mathcal{H}_0$ *and, if additionally* $n/k^{3/2} \to 0$ *holds, also under* $\mathcal{H}_1$

$$\widetilde{\sigma}^2 \xrightarrow[(n\to\infty)]{\mathrm{P}} \sigma^2.$$

Proof. Write

$$\begin{aligned}
\widetilde{\sigma}^2 &= \sum_{j=1}^{\lfloor n/r_n \rfloor} \left[\sum_{i=(j-1)r_n+1}^{jr_n} \frac{1}{\sqrt{k}} A_i(\mathbf{0}) - \frac{r_n}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{k}} A_i(\mathbf{0}) \right]^2 \\
&= \frac{1}{k} \sum_{j=1}^{\lfloor n/r_n \rfloor} \left[\sum_{i=(j-1)r_n+1}^{jr_n} \widetilde{A}_i(\mathbf{0}) - \frac{r_n}{n} \sum_{i=1}^{n} \widetilde{A}_i(\mathbf{0}) \right. \\
&\qquad \left. + \left(\frac{1}{\sqrt{k}} \sum_{i=(j-1)r_n+1}^{jr_n} g_n\left(\frac{i}{n}\right) - \frac{r_n}{n\sqrt{k}} \sum_{i=1}^{n} g_n\left(\frac{i}{n}\right) \right) \right]^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{k} \sum_{j=1}^{\lfloor n/r_n \rfloor} \left[\sum_{i=(j-1)r_n+1}^{jr_n} \widetilde{A}_i(\mathbf{0}) - \frac{r_n}{n} \sum_{i=1}^{n} \widetilde{A}_i(\mathbf{0}) \right]^2 \\
&\qquad + 2 \left[\sum_{i=(j-1)r_n+1}^{jr_n} \widetilde{A}_i(\mathbf{0}) - \frac{r_n}{n} \sum_{i=1}^{n} \widetilde{A}_i(\mathbf{0}) \right] \cdot \\
&\qquad\qquad \left[\frac{1}{\sqrt{k}} \sum_{i=(j-1)r_n+1}^{jr_n} g_n\left(\frac{i}{n}\right) - \frac{r_n}{n\sqrt{k}} \sum_{i=1}^{n} g_n\left(\frac{i}{n}\right) \right] \\
&\qquad + \left[\frac{1}{\sqrt{k}} \sum_{i=(j-1)r_n+1}^{jr_n} g_n\left(\frac{i}{n}\right) - \frac{r_n}{n\sqrt{k}} \sum_{i=1}^{n} g_n\left(\frac{i}{n}\right) \right]^2 \\
&=: (I) + (II) + (III).
\end{aligned}
$$

where we have assumed $r_n/n \in \mathbb{N}$ to avoid additional bookkeeping.

Note that by uniform boundedness of the g_n

$$
\left[\frac{1}{\sqrt{k}} \sum_{i=(j-1)r_n+1}^{jr_n} g_n\left(\frac{i}{n}\right) - \frac{r_n}{n\sqrt{k}} \sum_{i=1}^{n} g_n\left(\frac{i}{n}\right) \right] \leq 2 \frac{r_n}{\sqrt{k}} \max_{i \in \{1,\ldots,n\}} = \mathcal{O}\left(\frac{r_n}{\sqrt{k}}\right).
$$

Hence, by **(D1)** and $n/k^{3/2} \to 0$,

$$
|(III)| = \mathcal{O}\left(\frac{n}{r_n k} \left(\frac{r_n}{\sqrt{k}} \right)^2 \right) = o(1)
$$

and

$$
\begin{aligned}
|(II)| &\leq \mathcal{O}\left(\frac{r_n}{\sqrt{k}}\right) \frac{1}{k} \sum_{j=1}^{\lfloor n/r_n \rfloor} \left| \sum_{i=(j-1)r_n+1}^{jr_n} \widetilde{A}_i(\mathbf{0}) - \frac{r_n}{n} \sum_{i=1}^{n} \widetilde{A}_i(\mathbf{0}) \right| \\
&= \mathcal{O}\left(\frac{r_n}{\sqrt{k}}\right) \mathcal{O}_{\mathrm{P}}(1) = o_{\mathrm{P}}(1),
\end{aligned}
$$

because by Markov's inequality and Lemma 5.1

$$
\mathrm{P}\left\{ \frac{1}{k} \sum_{j=1}^{\lfloor n/r_n \rfloor} \left| \sum_{i=(j-1)r_n+1}^{jr_n} \widetilde{A}_i(\mathbf{0}) - \frac{r_n}{n} \sum_{i=1}^{n} \widetilde{A}_i(\mathbf{0}) \right| \geq K \right\}
$$

$$\leq K^{-1}\frac{1}{k}\sum_{j=1}^{\lfloor n/r_n\rfloor}\mathbb{E}\left|\sum_{i=(j-1)r_n+1}^{jr_n}\widetilde{A}_i(\mathbf{0})-\frac{r_n}{n}\sum_{i=1}^{n}\widetilde{A}_i(\mathbf{0})\right|=K^{-1}\mathcal{O}\left(\frac{n}{r_nk}\frac{r_nk}{n}\right)=K^{-1}\mathcal{O}(1).$$

(Under $\mathcal{H}_0$ we have $(II)=(III)=0$, such that the condition $n/k^{3/2}\to 0$ is not needed.)

Now decompose (I) as follows

$$(I)=\sum_{j=1}^{\lfloor n/r_n\rfloor}\left[\sum_{i=(j-1)r_n+1}^{jr_n}\frac{1}{\sqrt{k}}\widetilde{A}_i(\mathbf{0})\right]^2-\frac{r_n}{n}\left[\sum_{i=1}^{n}\frac{1}{\sqrt{k}}\widetilde{A}_i(\mathbf{0})\right]^2=:A_n-B_n \tag{5.39}$$

We have $B_n=o_{\mathrm{P}}(1)$ because of Lemma 5.2 and $r_n=o(n)$. Turning to A_n, let $\mathcal{O}_n$ denote the set of odd numbers in $1,\dots,\lfloor n/r_n\rfloor$. Then by **(C1)** and Eberlein (1984, Lem. 2 and the remarks below it)

$$\sum_{j\in\mathcal{O}_n}Y_{j,n}^2(\mathbf{0})\qquad\text{and}\qquad\sum_{j\in\mathcal{O}_n}\widetilde{Y}_{j,n}^2(\mathbf{0})$$

have the same limit distributions (if they exist), where $\widetilde{Y}_{j,n}^2(\mathbf{0})$ are independent copies of $Y_{j,n}^2(\mathbf{0})$. First note that by the c_r-inequality and Lemma 5.1

$$\begin{aligned}\frac{n}{r_n}\mathbb{E}\widetilde{Y}_{j,n}^4(\mathbf{0})&=\frac{n}{r_nk^2}\mathbb{E}\left[\sum_{i=(j-1)r_n+1}^{jr_n}\widetilde{A}_i(\mathbf{0})\right]^4\\&\leq\frac{n}{r_nk^2}r_n^3\sum_{i=(j-1)r_n+1}^{jr_n}\mathbb{E}\left[\widetilde{A}_i^4(\mathbf{0})\right]=\mathcal{O}\left(\frac{r_n^2}{k}\right)=o(1).\end{aligned}$$

Then use Tschebycheff's inequality to obtain

$$\begin{aligned}\mathrm{P}\left(\sum_{j\in\mathcal{O}_n}\left[\widetilde{Y}_{j,n}^2(\mathbf{0})-\mathbb{E}\widetilde{Y}_{j,n}^2(\mathbf{0})\right]\geq\varepsilon\right)&\leq\varepsilon^{-2}\sum_{j\in\mathcal{O}_n}\mathrm{Var}(\widetilde{Y}_{j,n}^2(\mathbf{0}))\\&\leq\varepsilon^{-2}\left|\mathcal{O}_n\right|\mathbb{E}\widetilde{Y}_{j,n}^4(\mathbf{0})=o(1).\end{aligned}$$

Combining this with

$$\sum_{j\in\mathcal{O}_n}\mathbb{E}\widetilde{Y}_{j,n}^2(\mathbf{0})=\frac{r_n}{n}\sum_{j\in\mathcal{O}_n}\frac{n}{r_nk}\mathrm{Var}\left(\sum_{i=(j-1)r_n+1}^{jr_n}I_{\{X_i>b_x,Y_i>b_y\}}\right)\overset{\textbf{(D2)}}{=}\sigma^2/2+o(1),$$

gives $\sum_{j\in\mathcal{O}_n}\widetilde{Y}_{j,n}^2(\mathbf{0})\to\sigma^2/2$ and hence $\sum_{j\in\mathcal{O}_n}Y_{j,n}^2(\mathbf{0})\to\sigma^2/2$ in probability. Similarly we get $\sum_{j\in\mathcal{O}_n}Y_{j,n}^2(\mathbf{0})\to\sigma^2/2$, where $\mathcal{E}_n:=1,\dots,\lfloor n/r_n\rfloor\setminus\mathcal{O}_n$. From these two convergences the result is now obvious. □

Now we show that the replacement of b_x, b_y by $X_{(k+1)}, Y_{(k+1)}$ can be done in the terms A_n and B_n defined in (5.39). The following lemma is for A_n.

Lemma 5.9. *Suppose assumptions* **(D1)** *and* **(D3)** *are met. Then*

$$\sup_{\boldsymbol{\xi}\in[-K,K]^2}\left|\sum_{j=1}^{\lfloor n/r_n\rfloor}\left[Y_{j,n}(\boldsymbol{\xi})^2-Y_{j,n}(\mathbf{0})^2\right]\right|=o_{\mathrm{P}}(1).$$

Proof. Expand

$$\begin{aligned}
&\sum_{j=1}^{\lfloor n/r_n\rfloor}\left[Y_{j,n}(\boldsymbol{\xi})^2-Y_{j,n}(\mathbf{0})^2\right]\\
=&\sum_{j=1}^{\lfloor n/r_n\rfloor}\Bigg[\frac{1}{k}\sum_{i,l=(j-1)r_n+1}^{jr_n}\{A_i(\mathbf{0})A_l(\mathbf{0})-A_i(\boldsymbol{\xi})A_l(\boldsymbol{\xi})\}\\
&\qquad-\frac{2}{k}\sum_{i,l=(j-1)r_n+1}^{jr_n}\{A_i(\mathbf{0})\mathbb{E}[A_l(\mathbf{0})]-A_i(\boldsymbol{\xi})\mathbb{E}[A_l(\boldsymbol{\xi})]\}\\
&\qquad+\frac{1}{k}\sum_{i,l=(j-1)r_n+1}^{jr_n}\{\mathbb{E}[A_i(\mathbf{0})]\mathbb{E}[A_l(\mathbf{0})]-\mathbb{E}[A_i(\boldsymbol{\xi})]\mathbb{E}[A_l(\boldsymbol{\xi})]\}\Bigg]\\
=:&(I)-(II)+(III).
\end{aligned}$$

In the following write $\boldsymbol{K}=(K,K)'$ for short. By Lemma 5.1 uniformly in $\boldsymbol{\xi}\in[-K,K]^2$

$$|(III)|=\mathcal{O}\left(\frac{r_nk}{n}\right)=o(1).$$

Next,

$$|(II)|=\left|\sum_{j=1}^{\lfloor n/r_n\rfloor}\frac{2}{k}\sum_{i,l=(j-1)r_n+1}^{jr_n}\Big\{A_i(\mathbf{0})\mathbb{E}[A_l(\mathbf{0})-A_l(\boldsymbol{\xi})]-\left[A_i(\boldsymbol{\xi})-A_i(\mathbf{0})\right]\mathbb{E}[A_l(\boldsymbol{\xi})]\Big\}\right|$$

$$
\begin{aligned}
&= \left\lfloor \frac{n}{r_n} \right\rfloor \frac{2}{k} r_n^2 \mathcal{O}\left(\frac{\sqrt{k}}{n}\right) + \left| \sum_{j=1}^{\lfloor n/r_n \rfloor} \frac{2}{k} \sum_{i,l=(j-1)r_n+1}^{jr_n} \left[A_i(\boldsymbol{\xi}) - A_i(\mathbf{0})\right] \mathbb{E}[A_l(\boldsymbol{\xi})] \right| \\
&= o(1) + \mathcal{O}\left(1\right) \left| \frac{r_n}{n} \sum_{j=1}^{\lfloor n/r_n \rfloor} \sum_{i=(j-1)r_n+1}^{jr_n} \left[A_i(\boldsymbol{\xi}) - A_i(\mathbf{0})\right] \right| \\
&\leq o(1) + \mathcal{O}(1) \left| \frac{r_n}{n} \sum_{j=1}^{\lfloor n/r_n \rfloor} \sum_{i=(j-1)r_n+1}^{jr_n} \left[A_i(-\boldsymbol{K}) - A_i(\boldsymbol{K})\right] \right|,
\end{aligned}
$$

where we used $A_i(\mathbf{0}) \leq 1$ and Lemma 5.3 for the second equality and Lemma 5.1 and **(D1)** for the third. Markov's inequality and Lemma 5.3 imply

$$
\begin{aligned}
&\mathrm{P}\left\{ \left| \frac{r_n}{n} \sum_{j=1}^{\lfloor n/r_n \rfloor} \sum_{i=(j-1)r_n+1}^{jr_n} \left[A_i(-\boldsymbol{K}) - A_i(\boldsymbol{K})\right] \right| \geq \varepsilon \right\} \\
\leq &\frac{1}{\varepsilon} \frac{r_n}{n} \left\lfloor \frac{n}{r_n} \right\rfloor r_n \max_{i \in \{1,\ldots,n\}} \mathbb{E}\left|A_i(-\boldsymbol{K}) - A_i(\boldsymbol{K})\right| = o(1).
\end{aligned}
$$

For the remaining term we get

$$
\left|(I)\right| \leq \sum_{j=1}^{\lfloor n/r_n \rfloor} \frac{1}{k} \sum_{i,l=(j-1)r_n+1}^{jr_n} \left|A_i(-\boldsymbol{K})A_l(-\boldsymbol{K}) - A_i(\boldsymbol{K})A_l(\boldsymbol{K})\right|.
$$

Then, using Markov's and the triangular inequality,

$$
\mathrm{P}\left\{\left|(I)\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon} \left\lfloor \frac{n}{r_n} \right\rfloor \frac{r_n^2}{k} \underbrace{\max_{i \in \{1,\ldots,n\}} \mathbb{E}\left|A_i(-\boldsymbol{K})A_l(-\boldsymbol{K}) - A_i(\boldsymbol{K})A_l(\boldsymbol{K})\right|}_{=\mathcal{O}(\sqrt{k}/n)} = o(1),
$$

where

$$
\mathbb{E}\left|A_i(-\boldsymbol{K})A_l(-\boldsymbol{K}) - A_i(\boldsymbol{K})A_l(\boldsymbol{K})\right| = \mathcal{O}(\sqrt{k}/n)
$$

follows similarly as in Lemma 5.3. □

The next lemma takes care of the replacement in B_n.

Lemma 5.10. *Suppose assumptions* **(D1)-(D3)** *are met. Then*

$$\sup_{\boldsymbol{\xi}\in[-K,K]^2}\left|\frac{r_n}{n}\left[\left(\sum_{i=1}^n \frac{1}{\sqrt{k}}\widetilde{A}_i(\mathbf{0})\right)^2 - \left(\sum_{i=1}^n \frac{1}{\sqrt{k}}\widetilde{A}_i(\boldsymbol{\xi})\right)^2\right]\right| = o_{\mathrm{P}}(1). \tag{5.40}$$

Proof. The left-hand side of (5.40) may be bounded by

$$\frac{r_n}{n}\left(\sum_{i=1}^n \frac{1}{\sqrt{k}}\widetilde{A}_i(\mathbf{0})\right)^2 + \sup_{\boldsymbol{\xi}\in[-K,K]^2}\left|\frac{r_n}{n}\left(\sum_{i=1}^n \frac{1}{\sqrt{k}}\widetilde{A}_i(\boldsymbol{\xi})\right)^2\right|.$$

We have $\sum_{i=1}^n \frac{1}{\sqrt{k}}\widetilde{A}_i(\mathbf{0}) = \mathcal{O}_{\mathrm{P}}(1)$ by Lemma 5.2 and

$$\begin{aligned}\sum_{i=1}^n \frac{1}{\sqrt{k}}\widetilde{A}_i(\mathbf{0}) &= \sum_{i=1}^n \frac{1}{\sqrt{k}}\left(\widetilde{A}_i(\boldsymbol{\xi}) - \widetilde{A}_i(\mathbf{0})\right) + \sum_{i=1}^n \frac{1}{\sqrt{k}}\widetilde{A}_i(\mathbf{0})\\ &= o_{\mathrm{P}}(1) + \mathcal{O}_{\mathrm{P}}(1) = \mathcal{O}_{\mathrm{P}}(1)\end{aligned}$$

uniformly on compact $\boldsymbol{\xi}$-sets by Lemma 5.4. As $r_n/n \to 0$, the result follows. □

Proof of Theorem 5.2: That the replacement can be made in A_n and B_n respectively defined in (5.39) can be seen similarly as for (5.38) using Lemmas 5.9 and 5.10 respectively instead of Lemma 5.6. □

Proof of Corollary 5.1: Use Lemma 5.2 and combine with Theorem 5.2. □

Proof of Theorem 5.3: The proof follows the lines of the proof of Theorem 5.1. The only difference is that instead of (5.37) we get

$$\sqrt{k}t(1-t)\left(\frac{1}{kt}\sum_{i=1}^{\lfloor nt\rfloor} A_i(\mathbf{0}) - \frac{1}{k(1-t)}\sum_{i=\lfloor nt\rfloor+1}^{n} A_i(\mathbf{0})\right) \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \sigma B(t) + C(t) \quad \text{in } D[0,1] \tag{5.41}$$

by properties of $g_n(\cdot)$. The rest of the proof goes through unaltered. □

6 Conclusion

In this dissertation we developed structural breaks tests for the extremal properties of time series. While the present dissertation is rather heavy on the theory side, our ultimate interest lay in providing tools for answering empirical questions. Applying our tests to 'real-life' data sets, we found no evidence for extreme quantile changes in WTI log-returns during the Iraqi invasion of Kuwait in 1990. However, using a sequential monitoring procedure, we found evidence for extreme quantile breaks in the log-returns of Bank of America stock during the recent financial crisis of 2007-08 without an accompanying break in the tail index. This suggests a break in scale that leaves the tail index unaffected.

Our tests may help in answering a wide set of deeper empirical questions: why do some crises not induce a change in the (extremal) properties of a financial time series (as the invasion of Kuwait in the WTI log-return application), while others do (the Asian crisis of 1997-98 investigated in Quintos *et al.*, 2001)? Why do tail index changes occur in some crises (again the Asian crisis) but not in others (Bank of America log-returns during the recent financial crisis investigated in Chapter 4)? What are the first signs of distress in financial markets - extremal distributional changes of returns or non-extremal distributional changes? The beginning of extremal co-movement of returns or increased serial extremal dependence?

Of course, our tests will not give direct answers to these questions. Yet, they may help in investigating past crises by identifying changes in the distribution and extremal dependence (if any occurred at all) which could then be a first step towards an explanation.

Bibliography

Aguilar M, Hill J. 2015. Robust score and portmanteau tests of volatility spillover. *Journal of Econometrics* **184**: 37–61.

Anderson T, Darling D. 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics* **23**: 193–212.

Andreou E, Ghysels E. 2009. Structural breaks in financial time series. In Andersen T, Davis R, Kreiß JP, Mikosch T (eds.) *Handbook of Financial Times Series*, Berlin: Springer, pages 839–870.

Artzner P, Delbaen F, Eber JM, Heath D. 1999. Coherent measures of risk. *Mathematical Finance* **9**: 203–228.

Asai M, McAleer M, Yu J. 2006. Multivariate stochastic volatility: A review. *Econometric Reviews* **25**: 145–175.

Aschersleben P, Wagner M, Wied D. 2015. Monitoring euro area real exchange rates. In Steland A, Rafajłowicz E, Szajowski K (eds.) *Springer Proceedings in Mathematics and Statistics 122 (12th Workshop on Stochastic Models, Statistics and Their Applications)*, New York: Springer, pages 363–370.

Aue A, Hörmann S, Horváth L, Hušková M, Steinebach J. 2012. Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory* **28**: 804–837.

Aue A, Hörmann S, Horváth L, Reimherr M. 2009. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37**: 4046–4087.

Baek C, Pipiras V, Wendt H, Abry P. 2009. Second order properties of distribution tails and estimation of tail exponents in random difference equations. *Extremes* **12**: 361–400.

Basrak B, Davis R, Mikosch T. 2002. Regular variation of GARCH processes. *Stochastic Processes and their Applications* **99**: 95–115.

Beirlant J, Goegebeur Y, Segers J, Teugels J. 2004. *Statistics of Extremes*. Chichester: Wiley.

Beirlant J, Vynckier P, Teugels J. 1996. Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association* **91**: 1659–1667.

Berkes I, Horváth L, Kokoszka P. 2003. Estimation of the maximal moment exponent of a GARCH(1,1) sequence. *Econometric Theory* **19**: 565–586.

Berkes I, Horváth L, Kokoszka P. 2004. Testing for parameter constancy in GARCH(p,q) models. *Statistics & Probability Letters* **70**: 263–273.

Bickel P, Wichura M. 1971. Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Statistics* **42**: 1656–1670.

Billingsley P. 1968. *Convergence of Probability Measures*. New York: Wiley.

Billingsley P. 1995. *Probability and Measure*. New York: Wiley.

Bingham N, Goldie C, Teugels J. 1987. *Regular Variation*. Cambridge: Cambridge Universtiy Press.

Bollerslev T, Chou R, Kroner K. 1992. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* **52**: 5–59.

Bradley R. 1986. Basic properties of strong mixing conditions. In Eberlein E, Taqqu M (eds.) *Dependence in Probability and Statistics*, Boston: Birkhäuser, pages 165–192.

Bradley R. 2007. *Introduction to strong mixing conditions Vol. I*. Heber City: Kendrick Press.

Breiman L. 1965. On some limit theorems similar to the arc-sine law. *Theory of Probability and its Applications* **10**: 351–360.

Bücher A. 2015. A note on weak convergence of the sequential multivariate empirical process under strong mixing. *Journal of Theoretical Probability* **28**: 1028–1037.

Bücher A, Jäschke S, Wied D. 2015. Nonparametric tests for constant tail dependence with an application to energy and finance. *Journal of Econometrics* **187**: 154–168.

Campbell R, Forbes C, Koedijk K, Kofman P. 2008. Increasing correlations or just fat tails? *Journal of Empirical Finance* **15**: 287–309.

Candelon B, Stratmans S. 2006. Testing for multiple regimes in the tail behavior of emerging currency returns. *Journal of International Money and Finance* **25**: 1187–1205.

Chan N, Li D, Peng L, Zhang R. 2013. Tail index of an AR(1) model with ARCH(1) errors. *Econometric Theory* **29**: 920–940.

Chavez-Demoulin V, Embrechts P, Sardy S. 2014. Extreme-quantile tracking for financial time series. *Journal of Econometrics* **181**: 44–52.

Chu JCS, Stinchcombe M, White H. 1996. Monitoring structural change. *Econometrica* **64**: 1045–1065.

Cramér H, Leadbetter M. 1967. *Stationary and Related Stochastic Processes.* New York: Wiley.

Csörgő S, Deheuvels P, Mason D. 1985. Kernel estimates of the tail index of a distribution. *The Annals of Statistics* **13**: 1050–1077.

Csörgő S, Viharos L. 1998. Estimating the tail index. In Szyszkowicz B (ed.) *Asymptotic Methods in Probability and Statistics*, Amsterdam: Elsevier, pages 833–881.

Dacarogna M, Gençay R, Müller U, Olsen R, Pictet O. 2001. *An Introduction to High-Frequency Finance.* San Diego: Academic Press.

Daníelsson J. 2011. *Financial Risk Forecasting.* Chichester: Wiley.

Daníelsson J, de Vries C. 1997. Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance* **4**: 241–257.

Daníelsson J, Jansen D, de Vries C. 1996. The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics - Theory and Methods* **25**: 711–720.

Daníelsson J, Jorgensen B, Samorodnitsky G, Sarma M, de Vries C. 2013. Fat tails, VaR and subadditivity. *Journal of Econometrics* **172**: 283–291.

Datta S, McCormick W. 1998. Inference for the tail parameter of a linear process with heavy tail innovations. *Annals of the Institute of Statistical Mathematics* **50**: 337–359.

Davidson J. 1994. *Stochastic Limit Theory.* Oxford: Oxford University Press.

Davis R, Mikosch T. 1998. The sample autocorrelatios of heavy-tailed processes with applications to ARCH. *The Annals of Statistics* **26**: 2049–2080.

Davis R, Mikosch T. 2009. The extremogram: A correlogram for extreme events. *Bernoulli* **15**: 977–1009.

Davis R, Mikosch T, Cribben I. 2012. Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics* **170**: 142–152.

de Haan L, Ferreira A. 2006. *Extreme Value Theory*. New York: Springer.

de Haan L, Peng L. 1998. Comparison of tail index estimators. *Statistica Neerlandica* **52**: 60–70.

de Haan L, Resnick S. 1987. On regular variation of probability densities. *Stochastic Processes and their Applications* **25**: 83–93.

Diebold F, Schuermann T, Stroughair J. 2000. Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance* **1**: 30–35.

Dierckx G, Teugels J. 2010. Change point analysis of extreme values. *Environmetrics* **21**: 661–686.

Drees H. 1998a. A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference* **66**: 95–112.

Drees H. 1998b. On smooth statistical tail functionals. *Scandinavian Journal of Statistics* **24**: 187–210.

Drees H. 2000. Weighted approximations of tail processes for β-mixing random variables. *The Annals of Applied Probability* **10**: 1274–1301.

Drees H. 2003. Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9**: 617–657.

Drees H. 2008. Some aspects of extreme value statistics under serial dependence. *Extremes* **11**: 35–53.

Drees H, Rootzén H. 2010. Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics* **38**: 2145–2186.

Dudley R. 1999. *Uniform Central Limit Theorems*. Cambridge: Cambridge University Press.

Eberlein E. 1984. Weak convergence of partial sums of absolutely regular sequences. *Statistics & Probability Letters* **2**: 291–293.

Einmahl J, de Haan L, Zhou C. 2016. Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B* **78**: 31–51.

Einmahl J, Gantner M, Sawitzki G. 2010. Asymptotics of the shorth plot. *Journal of Statistical Planning and Inference* **140**: 3003–3012.

Engle R. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007.

Fasen V, Klüppelberg C, Schlather M. 2010. High-level dependence in time series models. *Extremes* **13**: 1–33.

Frahm G, Junker M, Schmidt R. 2005. Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics & Economics* **37**: 80–100.

Gabaix X. 2009. Power laws in economics and finance. *Annual Review of Economics* **1**: 255–293.

Gaenssler P, Haeusler E. 1986. On martingale central limit theory. In Eberlein E, Taqqu M (eds.) *Dependence in Probability and Statistics*, Boston: Birkhäuser, pages 303–334.

Galbraith J, Zernov S. 2004. Circuit breakers and the tail index of equity returns. *Journal of Financial Econometrics* **2**: 109–129.

Geluk J, Peng L, De Vries C. 2000. Convolutions of heavy-tailed random variables and application to portfolio diversification and MA(1) time series. *Advances in Applied Probability* **32**: 1011–1026.

Gomes M, Pestana D. 2007. A sturdy reduced-bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association* **102**: 280–292.

Haeusler E, Teugels J. 1985. On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics* **13**: 743–756.

Hill B. 1975. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**: 1163–1174.

Hill J. 2009. On functional central limit theorems for dependent, heterogeneous arrays with applications to tail index and tail dependence estimation. *Journal of Statistical Planning and Inference* **139**: 2091–2110.

Hill J. 2010. On tail index estimation for dependent, heterogeneous data. *Econometric Theory* **26**: 1398–1436.

Hill J. 2011a. Extremal memory of stochastic volatility with an application to tail shape inference. *Journal of Statistical Planning and Inference* **141**: 663–676.

Hill J. 2011b. Tail and nontail memory with applications to extreme value and robust statistics. *Econometric Theory* **27**: 844–884.

Hsing T. 1991. On tail index estimation using dependent data. *The Annals of Statistics* **19**: 1547–1569.

Inclan C, Tiao G. 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* **89**: 913–923.

Janssen A, Drees H. 2016. A stochastic volatility model with flexible extremal dependence structure. *Bernoulli* **22**: 1448–1490.

Kelly B, Jiang H. 2014. Tail risk and asset prices. *Review of Financial Studies* **27**: 2841–2871.

Kim M, Lee S. 2009. Test for tail index change in stationary time series with Pareto-type marginal distribution. *Bernoulli* **15**: 325–356.

Kim M, Lee S. 2011. Change point test for tail index for dependent data. *Metrika* **74**: 297–311.

Kim M, Lee S. 2012. Change point test of tail index for autoregressive processes. *Journal of the Korean Statistical Society* **41**: 305–312.

Koedijk K, Schafgans M, de Vries C. 1990. The tail index of exchange rate returns. *Journal of International Economics* **29**: 93–108.

Kulik R, Soulier P. 2011. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications* **121**: 109–134.

Leadbetter M. 1983. Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **65**: 291–306.

Ledford A, Tawn J. 2003. Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **65**: 521–543.

Lin Z, Choi YK. 1999. Some limit theorems for fractional Lévy Brownian fields. *Stochastic Processes and their Applications* **82**: 229–244.

Lindner A. 2009. Stationarity, mixing, distributional properties and moments of GARCH(p,q)-processes. In Andersen T, Davis R, Kreiß JP, Mikosch T (eds.) *Handbook of Financial Times Series*, Berlin: Springer, pages 43–70.

Longin F. 2000. From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance* **24**: 1097–1130.

Mandelbrot B. 1963. The variation of certain speculative prices. *The Journal of Business* **36**: 394–419.

Marimoutou V, Raggad B, Trabelso A. 2009. Extreme value theory and value at risk: Application to oil market. *Energy Economics* **31**: 519–530.

Mikosch T. 2005. How to model multivariate extremes if one must? *Statistica Neerlandica* **59**: 324–338.

Móricz F. 1982. A general moment inequality for the maximum of partial sums of single series. *Acta Scientiarum Mathematicarum* **44**: 67–75.

Quintos C, Fan Z, Phillips P. 2001. Structural change tests in tail behaviour and the Asian crisis. *Review of Economic Studies* **68**: 633–663.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Resnick S. 2007. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.

Rootzén H. 2009. Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications* **119**: 468–490.

Rothe C, Wied D. 2013. Misspecification testing in a class of conditional distributional models. *Journal of the American Statistical Association* **108**: 314–324.

Shao X, Zhang X. 2010. Testing for change points in time series. *Journal of the American Statistical Association* **105**: 1228–1240.

Shorack G, Wellner J. 1996. *Weak Convergence and Empirical Processes*. New York: Springer.

Stărică C, Granger C. 2005. Nonstationarities in stock returns. *The Review of Economics and Statistics* **87**: 503–522.

Sun P, Zhou C. 2014. Diagnosing the distribution of GARCH innovations. *Journal of Empirical Finance* **29**: 287–303.

van der Vaart A, Wellner J. 1996. *Weak Convergence and Empirical Processes*. New York: Springer.

Vogelsang T. 1997. Wald-type tests for detecting breaks in the trend function of a dynamic time series. *Econometric Theory* **13**: 818–849.

Vogelsang T. 1999. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics* **88**: 283–299.

Wagner N, Marsh T. 2004. Tail index estimation in small samples: Simulation results for independent and ARCH-type financial return models. *Statistical Papers* **45**: 545–562.

Weissman I. 1978. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association* **73**: 812–815.

Werner T, Upper C. 2004. Time variation in the tail behavior of bund future returns. *Journal of Future Markets* **24**: 387–398.

Wichura M. 1970. On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *The Annals of Mathematical Statistics* **41**: 284–291.

Wied D, Galeano P. 2013. Monitoring correlation change in a sequence of random variables. *Journal of Statistical Planning and Inference* **143**: 186–196.

Wied D, Krämer W, Dehling H. 2012. Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory* **28**: 570–589.

Xu K. 2015. Testing for structural change under non-stationary variances. *Econometrics Journal* **18**: 274–305.

Zeileis A, Leisch F, Kleiber C, Hornik K. 2005. Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics* **20**: 99–121.

Eidesstattliche Erklärung

Hiermit versichere ich, diese Arbeit selbstständig verfasst zu haben, und dabei keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Zitate wurden an den entsprechenden Stellen in der Arbeit kenntlich gemacht.

Leverkusen, 02.05.2016

Yannick Peer Hoga